

36:3 BERKELEY TECHNOLOGY LAW JOURNAL

2021

**Pages
861
to
1340**

Berkeley Technology Law Journal
Volume 36, Number 3

Production: Produced by members of the *Berkeley Technology Law Journal*.
All editing and layout done using Microsoft Word.

Printer: Joe Christensen, Inc., Lincoln, Nebraska.

Printed in the U.S.A.

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Library Materials, ANSI Z39.48—1984.

Copyright © 2021 Regents of the University of California.

All Rights Reserved.

Berkeley Technology Law Journal

University of California

School of Law

3 Law Building

Berkeley, California 94720-7200

editor@btlj.org

<https://www.btlj.org>



BERKELEY TECHNOLOGY LAW JOURNAL

VOLUME 36

NUMBER 3

2021

TABLE OF CONTENTS

ARTICLES

FOREWORD LEX REFORMATICA: FIVE PRINCIPLES OF POLICY REFORM FOR THE TECHNOLOGICAL AGE.....	861
<i>Sonia K. Katyal</i>	
TECHNOLOGICAL “DISRUPTION” OF THE LAW’S IMAGINED SCENE: SOME LESSONS FROM <i>LEX INFORMATICA</i>	883
<i>Margot E. Kaminski</i>	
LEX AI: REVISITING PRIVATE ORDERING BY DESIGN.....	915
<i>Niva Elkin-Koren & Karni A. Chagal-Feferkorn</i>	
TECHNOLOGY LAW AS A VEHICLE FOR TECHNOLOGY JUSTICE: STOP ISP THROTTLING TO PROMOTE DIGITAL EQUITY.....	963
<i>Catherine J.K. Sandoval</i>	
FROM LEX INFORMATICA TO THE CONTROL REVOLUTION.....	1017
<i>Julie E. Cohen</i>	
RACIAL SEGREGATION AND THE DATA-DRIVEN SOCIETY: HOW OUR FAILURE TO RECKON WITH ROOT CAUSES PERPETUATES SEPARATE AND UNEQUAL REALITIES.....	1051
<i>Rashida Richardson</i>	
ALLOCATING RESPONSIBILITY IN CONTENT MODERATION: A FUNCTIONAL FRAMEWORK.....	1091
<i>Deirdre K. Mulligan & Kenneth A. Bamberger</i>	
AUTOMATED VIDEO INTERVIEWING AS THE NEW PHRENOLOGY.....	1173
<i>Ifeoma Ajunwa</i>	
REVISITING ROOMMATES.COM.....	1227
<i>G.S. Hans</i>	
A TRIBUTE TO JOEL REIDENBERG.....	1253
<i>Paul M. Schwartz</i>	
PRIVACY AS/AND CIVIL RIGHTS.....	1265
<i>Tiffany C. Li</i>	

CONTENT MODERATION AS SURVEILLANCE 1297

Hannah Bloch-Webba

SUBSCRIBER INFORMATION

The *Berkeley Technology Law Journal* (ISSN1086-3818), a continuation of the *High Technology Law Journal* effective Volume 11, is edited by the students of the University of California, Berkeley, School of Law and is published in print three times each year (March, September, December), with a fourth issue published online only (July), by the Regents of the University of California, Berkeley. Periodicals Postage Rate Paid at Berkeley, CA 94704-9998, and at additional mailing offices. POSTMASTER: Send address changes to Journal Publications, University of California, Berkeley Law—Library, LL123 Boalt Hall—South Addition, Berkeley, CA 94720-7210.

Correspondence. Address all correspondence regarding subscriptions, address changes, claims for non-receipt, single copies, advertising, and permission to reprint to Journal Publications, University of California, Berkeley Law—Library, LL123 Boalt Hall—South Addition, Berkeley, CA 94705-7210; (510) 643-6600; JournalPublications@law.berkeley.edu. *Authors:* see section titled Information for Authors.

Subscriptions. Annual subscriptions are \$65.00 for individuals and \$85.00 for organizations. Single issues are \$30.00. Please allow two months for receipt of the first issue. Payment may be made by check, international money order, or credit card (MasterCard/Visa). Domestic claims for non-receipt of issues should be made within 90 days of the month of publication; overseas claims should be made within 180 days. Thereafter, the regular back issue rate (\$30.00) will be charged for replacement. Overseas delivery is not guaranteed.

Form. The text and citations in the *Journal* conform generally to the THE CHICAGO MANUAL OF STYLE (16th ed. 2010) and to THE BLUEBOOK: A UNIFORM SYSTEM OF CITATION (Columbia Law Review Ass'n et al. eds., 20th ed. 2015). Please cite this issue of the *Berkeley Technology Law Journal* as 36 BERKELEY TECH. L.J. 3 (2021).

BTLJ ONLINE

The full text and abstracts of many previously published *Berkeley Technology Law Journal* articles can be found at <https://www.btlj.org>. Our site also contains a cumulative index; general information about the *Journal*; the BTLJ Blog, a collection of short comments and updates about new developments in law and technology written by BTLJ members; and *BTLJ Commentaries*, an exclusively online publication for pieces that are especially time-sensitive and shorter than typical law review articles.

INFORMATION FOR AUTHORS

The Editorial Board of the *Berkeley Technology Law Journal* invites the submission of unsolicited manuscripts. Submissions may include previously unpublished articles, essays, book reviews, case notes, or comments concerning any aspect of the relationship between technology and the law. If any portion of a manuscript has been previously published, the author should so indicate.

Format. Submissions are accepted in electronic format through Scholastica online submission system. Authors should include a curriculum vitae and resume when submitting articles, including his or her full name, credentials, degrees earned, academic or professional affiliations, and citations to all previously published legal articles. The Scholastica submission website can be found at <https://btlj.scholasticahq.com/for-authors>.

Citations. All citations should conform to THE BLUEBOOK: A UNIFORM SYSTEM OF CITATION (Columbia Law Review Ass'n et al. eds., 21st ed. 2020).

Copyrighted Material. If a manuscript contains any copyrighted table, chart, graph, illustration, photograph, or more than eight lines of text, the author must obtain written permission from the copyright holder for use of the material.

DONORS

The *Berkeley Technology Law Journal* and the Berkeley Center for Law & Technology acknowledge the following generous donors to Berkeley Law's Law and Technology Program:

Partners

WHITE & CASE LLP

Benefactors

BAKER BOTTS LLP

MORRISON & FOERSTER LLP

COVINGTON & BURLING LLP

ORRICK HERRINGTON & SUTCLIFFE
LLP

FENWICK & WEST LLP

PAUL HASTINGS LLP

FISH & RICHARDSON P.C.

SIDLEY AUSTIN LLP

KIRKLAND & ELLIS LLP

WEIL, GOTSHAL & MANGES LLP

LATHAM & WATKINS LLP

WILMER CUTLER PICKERING HALE
AND DORR LLP

MCDERMOTT WILL & EMERY LLP

WILSON SONSINI GOODRICH &
ROSATI

WINSTON & STRAWN LLP

Corporate, Government, Individual, and Foundation Sponsors

ATLASSIAN	LITINOMICS, INC.
BOIES SCHILLER & FLEXNER LLP	MARKS & CLERK LLP
CISCO SYSTEMS, INC.	MICROSOFT CORPORATION
THE CITIZEN LAB	MOZILLA CORPORATION
COMCAST CABLE	NOKIA CORPORATION
CORNERSTONE RESEARCH	PALANTIR TECHNOLOGIES
DARTS IP	QUALCOMM INCORPORATED
GEN LAW FIRM	RLM TRIALGRAPHIX
GOODWIN PROCTER LLP	STARZ
GOOGLE INC.	TYSON & MENDES
INTEL CORPORATION	UNIFY CONSULTING
INVENTIONSHARE INC.	VIA LICENSING CORPORATION
JENNER & BLOCK	VYNL
KILBURN & STRODE	WESTERN DIGITAL

Members

BAKER & MCKENZIE LLP

KILPATRICK TOWNSEND &
STOCKTON LLP

BEIJING EAST IP

KNOBBE MARTENS LLP

DESMARAIS LLP

MORGAN LEWIS & BROCKIUS

DURIE TANGRI LLP

ROBINS KAPLAN, MILLER & CIRESI
LLP

GREENBERG TRAURIG LLP

TENSEGRITY LAW GROUP LLP

GTC LAW GROUP LLP & AFFILIATES

VAN PELT, YI & JAMES LLP

HAYNES AND BOONE, LLP

WANHUIDA INTELLECTUAL
PROPERTY

IRELL & MANELLA LLP

WILLKIE FARR & GALLAGHER LLP

KEKER VAN NEST & PETERS LLP

WOMBLE BOND DICKINSON LLP

BOARD OF EDITORS

2020–2021

Executive Board

Editor-in-Chief
S. EMMA LEE

Senior Articles Editors
MUHTADI CHOUDHURI
MATT CHUNG
MARTA ROCHA

Senior Executive Editor
HARRISON GERON

Senior Production Editor
HAILEY YOOK

Managing Editor
MADISON BOWER

Senior Scholarship Editor
ANGELA GRIGGS

Senior Student Publication Editors
WALTER MOSTOWY
KEVIN YANG

Senior Online Content Editor
ALLAN E. HOLDER

Editorial Board

Submissions Editors
JOHN BATOHA
GRACE MCFEE
THOMAS HORN

Production Editors
ROBIN CHANG
JOELLE FERGUSON
EMILY ROBERTS

Technical Editors
SALLY CHOI
MIN JUNG “MJ” HAN
ALEX HARVEY
JOSEPH KINGERSKI

Student Publication Editors
JENNIFER CHUNG
ANUJ EZEKIEL

Notes & Comments Editors
LOC HO
SHREYA SANTHANAM

Symposium Editors
MARGARET LYNCH
DEBBIE MOSLEY

Web & Technology Editors
KARNIK HAJJAR
HENRY METRO

Podcast Editors
HALEY BROUGHTON
ANDY ZACHRICH

LLM Editor
MARIO MARTINEZ

Member Relations Editor
RACHEL WILSON

Alumni Relations Editor
ARMBIEN SABILLO

External Relations Editor
GRACE (HJ) KIM

Commentaries Editor
VERONICA BOGNOT

Articles Editors
LIAM AZARTASH
KEVIN CHEN
NATALIE T. CRAWFORD
JAMESON DAVIS
JASON FRANCIS

Articles Editors
JEFFREY JACOBSEN
TOM JAMES
JOSEPH KROON
CHARLIZE MORGAN
ALEX MCKENZIE
SHALEV NETANEL

Articles Editors
YEMAJ SHEIK
DAKOTA SNEED
BLAINE VALENCIA
SOPHIA WALLACH
ALI ZARRABI

MEMBERSHIP

Vol. 36 No. 3

Associate Editors

JONATHAN BAER	KAVYA DASARI	FATIMA LADHA
BOGDAN BELEI	KURT FREDERICKSON	WYATT LARKIN
SETH BERTOLUCCI	RAFI GINSBURG	MATT SARDO
CONNOR BOEHM	REBECCA HO	SARA TSAI
HALEY BROUGHTON	NATHANIEL KELLERER	JESSICA WANG
LUCILLE DAI-HE	CONNOR KENNEDY	

Members

JARED ABES	EDUARDO FIGUEROA	ROSS MOODY
RICH ABIDOR	COLE GINGRICH	ERIN MOORE
IAN AFLAGUE	KHASH GOSHTASBI	GAYATRI PARANJAPE
TIFFANY ALLEN	SAVANNAH GROSSARTH	JUSTINE MCCARTHY POTTER
SCOTT ARONIN	DYLAN HOULE	BREANNA QIN
PIERRE BARTHELEMY	CHRISTINA JOHNSON	JENNY QUANG
MIKE BEHLEN	NATALIE KALISS	MEIRAM RAKHIMBEKOV
BROOKE D'AMORE BRADLEY	ROGER KANG	LIZ FREEMAN ROSENZWEIG
HANNAH BROWN	CHRISTIAN KNIPFER	BARBARA ROWINSKA
ZHIWEI CAI	PAULINE LE	CATIE SAKURAI
KEVIN CHIU	DIANA LEE	WILL SERIO
MACKENZIE CONCEPCION	GARY LEE	EVA SPITZEN
ALEXA DAUGHERTY	VALENTINO LUCINI	MEGHAN SULLIVAN
VICTORIA DIPLA	MATTHEW LUEVANO	JENNIFER SUN
MADELINE ELKINS	MAHAA MAHMOOD	RACHEL THOMPSON
CLINTON EWELL	BLAINE MANIRE	MICHELLE WONG
ROBERT FAIRBANKS	THOMAS MATTES	DIMING XU
CITRA FATIHAH	ISHITA MATTOO	ANGELA ZHAO
KAYLA FEDLER	ALISTAIR MCINTYRE	MICHELLE ZIPERSTEIN
	DANIEL METEER	

BTLJ ADVISORY BOARD

JIM DEMPSEY
*Executive Director of the
Berkeley Center for Law & Technology*
U.C. Berkeley School of Law

ROBERT C. BERRING, JR.
Walter Perry Johnson Professor of Law, Emeritus
U.C. Berkeley School of Law

MATTHEW D. POWERS
Tensegrity Law Group, LLP

JESSE H. CHOPER
Earl Warren Professor of Public Law
U.C. Berkeley School of Law

PAMELA SAMUELSON
*Richard M. Sherman Distinguished Professor of
Law & Information and Faculty Director of the
Berkeley Center for Law & Technology*
U.C. Berkeley School of Law

REGIS MCKENNA
Chairman and CEO
Regis McKenna, Inc.

LIONEL S. SOBEL
*Professor of Law, Emeritus and Director of the
International Entertainment & Media Law
Summer Program in London*
Southwestern University School of Law

PETER S. MENELL
*Koret Professor of Law and Faculty
Director of the Berkeley Center
for Law & Technology*
U.C. Berkeley School of Law

LARRY W. SONSINI
Wilson Sonsini Goodrich & Rosati

ROBERT P. MERGES
*Wilson Sonsini Goodrich & Rosati Professor of
Law and Faculty
Director of the Berkeley Center
for Law & Technology*
U.C. Berkeley School of Law

MICHAEL STERN
Cooley LLP

DEIRDRE K. MULLIGAN
*Assistant Professor and Faculty Director of the
Berkeley Center for Law and Technology*
U.C. Berkeley School of Information

MICHAEL TRAYNOR
Cobalt LLP

JAMES POOLEY
James Pooley, PLC

THOMAS F. VILLENEUVE
Gunderson Dettmer Stough Villeneuve
Franklin & Hachigian LLP

BERKELEY CENTER FOR LAW & TECHNOLOGY 2020–2021

Executive Director

JIM DEMPSEY

Faculty Directors

KENNETH A. BAMBERGER	PETER S. MENELL	PAUL SCHWARTZ
CATHERINE CRUMP	ROBERT P. MERGES	ERIK STALLMAN
CATHERINE FISK	DEIRDRE K. MULLIGAN	JENNIFER M. URBAN
CHRIS HOOFNAGLE	TEJAS N. NARECHANIA	MOLLY S. VAN HOUWELING
SONIA KATYAL	ANDREA ROTH	REBECCA WEXLER
ORIN KERR	PAMELA SAMUELSON	

Fellow

KATHRYN HASHIMOTO	YUAN HAO
-------------------	----------

Staff

MARK COHEN	RICHARD RISK
NATALIE COLETTA	MATTHEW RAY
JANN DUDLEY	IRYS SCHENKER

FOREWORD

LEX REFORMATICA: FIVE PRINCIPLES OF POLICY REFORM FOR THE TECHNOLOGICAL AGE

Sonia K. Katyal[†]

I. INTRODUCTION

Almost twenty five years ago, our beloved former colleague Joel Reidenberg penned an article that argued that law and government regulation were not the only source of authority and rulemaking in the Information Society.¹ Rather, he argued that technology itself, particularly system design choices like network design and system configurations, can also impose similar regulatory norms on communities.² These rules and systems, he argued, comprised a Lex Informatica—a term that Reidenberg coined in historical reference to “Lex Mercatoria,” a system of international, merchant-driven norms in the Middle Ages that emerged independent of localized sovereign control.

This work was an iconic piece of literature. For Reidenberg, there were clear parallels between merchants’ travels across borders and today’s users, traveling across the information superhighway, surpassing local sovereignty and confronting a tangle of conflicting regulations along the way.³ It is no surprise that this landmark article was published in 1998, the same year that Congress passed the Digital Millennium Copyright Act (DMCA), one of the first attempts of legislators to address the onset of the digital era.⁴ And Reidenberg’s theory provided the backbone for yet another tour de force that defined the relationship between law and the internet, Lawrence Lessig’s *Code and Other Laws of Cyberspace*, published shortly afterward.⁵

DOI: <https://doi.org/10.15779/Z38HD7NT5B>

© 2022 Sonia K. Katyal.

† Associate Dean for Faculty Development and Research and Haas Distinguished Professor, University of California at Berkeley School of Law. The author wishes to thank each of the wonderful authors who wrote papers for their insights and conversation, Pamela Samuelson, Paul Schwartz, Rebecca Wexler, and Angela Zhao, who offered excellent research assistance and support.

1. Joel R. Reidenberg, *Lex Informatica: The Formulation of Information Policy Rules Through Technology*, 76 Tex. L. Rev. 553, 554 (1998).

2. *Id.* at 554–55.

3. *Id.* at 554.

4. See Digital Millennium Copyright Act, H.R. 2281, 105th Cong. § 6 (1997).

5. See generally LAWRENCE LESSIG, CODE AND OTHER LAWS OF CYBERSPACE (1999).

Today, however, we confront a different phenomenon, one that requires us to draw upon the wisdom of Reidenberg's landmark work in considering the repercussions of the previous era. As much as Lex Informatica provided us with a descriptive lens to analyze the birth of the internet, we are now confronted with the aftereffects of decades of muted, if not absent, regulation. When technological social norms are allowed to develop outside of clear legal restraints, who wins? Who loses? In this new era, we face a new set of challenges—challenges that force us to confront a critical need for infrastructural reform that focuses on the interplay between public and private forms of regulation (and self-regulation), its costs, and its benefits.

This turn is undeniably significant, and it draws similarities to an earlier period in intellectual property history. In the 1980s and 1990s, a body of scholarship began to flourish that emphasized postmodern, feminist, critical race, and post-colonial approaches to intellectual property and information law.⁶ These approaches, in part, focused their gaze towards how intellectual property entitlements facilitated the widening scope of inequality for some and also served as a source of empowerment for others.⁷ Other scholars studied how even commons-like frameworks, seemingly open and free for all to use, actually benefited the powerful at the cost of disenfranchised groups. Madhavi Sunder and Anupam Chander, for example, authored a milestone article that pointed out how the concept of a public domain—and the romance attached to it—unwittingly contributes to a widening scope of inequality by offering a racialized libertarianism that celebrates appropriation, often at the cost of

6. Carys J. Craig, *Critical Copyright Law and the Politics of 'IP'* in RESEARCH HANDBOOK ON CRITICAL LEGAL THEORY 301, 304 (Emilios Christodoulidis, Ruth Dukes & Marco Goldoni eds., 2019).

7. *Id.*

8. See Boatema Boetang, *Symposium: Walking the Tradition-Modernity Tightrope: Gender Contradictions in Textile Production and Intellectual Property Law in Ghana*, 15 AM. U. J. GENDER SOC. POL'Y & L. 341, 345–46 (2007) (noting how cultural products of indigenous peoples are appropriated from the public domain and then repackaged as protected intellectual property, such as Ghanaian cloth designs); RUTH L. OKEDJI, CTR. INT'L GOVERNANCE INNOVATION, TRADITIONAL KNOWLEDGE AND THE PUBLIC DOMAIN 15–16 (2018) (arguing that the public domain benefits existing beneficiaries of the IP system and undermines creativity and innovation in local communities and indigenous peoples); K.J. Greene, *Intellectual Property at the Intersection of Race and Gender: Lady Sings the Blues*, 16 AM. U. J. GENDER, SOC. POL'Y & L. 365, 370–71 (noting how copyright law appropriated Black cultural production and made invisible the contributions of Black ragtime, blues, and jazz artists); Anjali Vats & Deidré A. Keller, *Critical Race IP*, 36 CARDOZO ARTS & ENT. L.J. 735 (2018) (outlining a scholarly movement that focuses on the racial disparities generated by the enforcement and ownership of intellectual property).

marginalized groups.⁸ In addition, these works were also linked to the emerging field of Critical Information Studies (CIS), a field that focused on studying the abilities, rights, and limitations of the ways that users, consumers, or citizens alter or critique cultural texts, and the role of property rights and other forms of information control in limiting flows of information.⁹

These earlier works, I would argue, have attained even greater salience in contemporary times, where the Black Lives Matter movement rightfully forces us to reckon with the need for a deeper interrogation—or perhaps integration—of the frameworks of social justice and intellectual property.¹⁰ Instead of demonstrating the richness, complexity, and promise of yesterday’s internet age, today’s events show us what precisely can happen in an age of information libertarianism, underscoring the need for a new approach to information regulation. The articles in this Issue are taken from two separate symposiums—one on Lex Informatica and another on race and technology law. At present, a conversation between them could not be any more necessary. Taken together, these papers showcase what I refer to as the Lex Reformatica of today’s digital age. This collection of papers demonstrates the need for scholars, lawyers, and legislators to return to Reidenberg’s foundational work and to update its trajectory towards a new era that focuses on the design of a new approach to reform.

Below, I highlight five principle themes drawn from these collected articles that showcase the need for new generations of reform and regulation, i.e., what I refer to as today’s ‘Lex Reformatica.’ The first concerns the need for infrastructural reform; the second involves a close attention to the negative impacts of underregulation; the third involves a focus on design and, relatedly, the concept of design justice; the fourth on reforming the interplay among public and private forms of regulation; and the final principle, which emphasizes the value of ex ante, instead of ex post forms of remediation.

9. Joel R. Reidenberg, *Lex Informatica: The Formulation of Information Policy Rules Through Technology*, 76 Tex. L. Rev. 553, 554 (1998). See Siva Vaidhyanathan, Afterword: Critical Information Studies, A Bibliographic Manifesto, 20 Cultural Stud. 292, 293 (2006).

10. For more commentary on the intersection between intellectual property and social justice, see Lateef Mtima, *IP Social Justice Theory: Access, Inclusion and Empowerment*, 55 GONZ. L. REV. 401 (2020); HANDBOOK OF INTELLECTUAL PROPERTY AND SOCIAL JUSTICE (Stephen Jamar & Lateef Mtima eds., 2021); Peter Menell, *Property, Intellectual Property and Social Justice: Mapping the Next Frontier*, 5 BRIGHAM-KANNER PROP. RTS. CONF. J. 147 (2016); and Anupam Chander & Madhavi Sunder, *Is Nozick Kicking Rawls’s Ass? Intellectual Property and Social Justice*, 40 U.C. Davis L. Rev. 563 (2007), along with the collection of essays in the Symposium.

II. THE NEED FOR INFRASTRUCTURAL REFORM

At the heart of Reidenberg's insight was an analogy between the newly networked environment and the instability that early merchants faced in navigating differing jurisdictions.¹¹ For him, the management of content, personal information, and the preservation of ownership were three core areas of disruption. To navigate the uncertainties surrounding each domain, Reidenberg drew on early principles of *Lex Mercatoria* to argue that parties can carve out their own set of customs and practices, independent of local rules but which assured "basic fairness in their relationship."¹² Applying these principles to information technology, Reidenberg contended that information rules, i.e., rules of design, could do the same thing.

As several of these papers describe, *Lex Informatica* inspired a host of works that studied the norms, customs, and practices that characterized the growth of the internet, a world that emerged, initially, largely free from direct, legal regulation (and one that was considered superior for precisely this reason). Scholars were deeply skeptical of the need for or the benefits of regulation. Often, any discussion of regulation was immediately equated with overregulation. Consider one representative observation from this period from Lawrence Lessig: "Overregulation," he wrote, "stifles creativity. It smothers innovation. It gives dinosaurs a veto over the future. It wastes the extraordinary opportunity for a democratic creativity that digital technology enables."¹³

This view, widely shared by internet law academics at the time, aptly characterized the initial generation of internet-related scholarship, barely disguising a distrust of regulation and state overreach. In Lessig's single quote, we see, essentially, the union of three presumptive ideals in digital technology—first, the idea that regulation would stifle innovation; second, implicitly, that digital technology ruled by norms, rather than law, was preferable; and third, that digital technology enabled a democratization of creativity.

In a sense, *Lex Informatica* typified, and inspired this view. Yet if the initial growth of the Web's infrastructure might be characterized by the absence of direct regulation, law still played an indirectly powerful role in underscoring the trajectory of the Web's initial growth. And this is where our opening essay,

11. Reidenberg, *supra* note 1, at 55–54.

12. *Id.* at 553.

13. LAWRENCE LESSIG, *FREE CULTURE: HOW BIG MEDIA USES TECHNOLOGY AND THE LAW TO LOCK DOWN CULTURE AND CONTROL CREATIVITY* 151 (2004).

from Rashida Richardson, becomes most salient. As Richardson has astutely observed, law does not exist in a vacuum—its background influence has contributed to a myriad of issues surrounding digital inequality.¹⁴ These inequalities are intimately linked, both to the presence and absence of legal regulation. Consider, as she points out, the ways in which the history of de jure and de facto segregation in the United States has fed into the assembly and collection of training data, feeding algorithmic systems that generate AI-driven products that perpetuate racial inequality.¹⁵

Taking Richardson's work one step further and viewing it in another light, one can see how her powerful observations about the legacy of racial segregation requires us to think on an infrastructural level about the need for a critical analysis of data-driven technologies and the products that they create. As she astutely points out, segregation can reassert itself in a myriad of local, contemporary formations involving structural inequality.¹⁶ This requires a closer examination of specific historical and contemporary laws, customs and social practices (norms) to see how bias can reassert itself.¹⁷ And this interrogation, Richardson suggests, must take place at a local level.¹⁸ Here, Richardson argues that a focus on technological injustice is incomplete without studying the historical practices that contribute to systemic inequality:

These fields [focusing on legal procedures to create greater transparency or oversight] and the interventions they produce generally have two issues. First, they fail to reckon with the disadvantages and harms that preceded and are often compounded by data driven interventions. Second, they fail to decenter technology as the primary lens of analysis or modality of prevention and redress.¹⁹

The legal solutions to bias that are often promulgated, she points out, are mostly procedural in nature, she points out, and thus risk entrenching structural inequality.²⁰

Threaded throughout Richardson's powerful article is the idea that in order to address algorithmic bias, we need to study the historical infrastructure that

14. Rashida Richardson, *Racial Segregation and the Data-Driven Society: How Our Failure to Reckon with Root Causes Perpetuates Separate and Unequal Realities*, 36 BERKELEY TECH. L.J. 1051, 1053 (2021).

15. *Id.* at 104–6.

16. *Id.* at 106.

17. *Id.*

18. *Id.* at 108 (calling for clarity on the nature of the problem locally).

19. *Id.* at 135–36.

20. *Id.* at 136.

promotes structural inequalities. Only by decentering technology as the main lens of analysis and looking to the root cause of injustice, Richardson notes, can we begin to address algorithmic bias and other data-driven technologies that engender inequality. Here, too, Richardson offers us an intervention that focuses on infrastructural change: transformative justice. Her conclusion closes with a call towards employing a transformative justice framework, which she argues uses a systems-oriented approach that examines collective societal responsibility in creating systemic harms and centers people who are often “excluded from but pivotal to” the issues that data driven technologies raise.²¹ Transformative justice, she argues, is necessary to create meaningful interventions to the problems of data-driven technology and advance society beyond the status quo.²²

Richardson’s call to excavate the infrastructure of historical inequality and transformative justice is nicely mirrored by Sandoval’s excellent paper on ISP throttling, which also evokes a similar concern for infrastructural justice, or (as she calls it) “technology justice.”²³ Whereas Richardson focuses on the historical framing of segregation and its contributions to algorithmic bias, Sandoval offers us a contemporary illustration of how structural and historical inequality can fuel a deprivation of access to critically important information. As she argues, “[i]nfrastructure regulation creates the future’s physical and social architecture,” pointing out that slowing down of access to the internet leaves users unable to access news sources, telemedicine, or to use videoconferencing necessary for work or school.²⁴

While Richardson calls for a framing that focuses on transformative justice, Sandoval’s elucidation of technology justice offers a similar, complementary reframing. She draws from the historical roots of the digital divide, showing us its contemporary aftereffects in the problem of ISP throttling. But even more presciently, she offers us a solution that focuses both on reforming the notion of transparency and the notion of internet access simultaneously:

Inadequate disclosure in small faded print that does not make the consequences of ISP throttling clear is inconsistent with internet openness and may violate FCC transparency and FTC deceptive

21. *Id.* at 139–40.

22. *Id.* at 140.

23. See generally Catherine J.K. Sandoval, *Technology Law As A Vehicle For Technology Justice: Stop ISP Throttling To Promote Digital Equity*, 36 BERKELEY TECH. L.J. 963 (2021) (referring to “technology justice” in the title and various parts of the article).

24. *Id.* at 983.

conduct laws and regulations. The prevalence of inadequate disclosures across carriers underscores the need for FCC and FTC regulatory action to protect consumers, internet openness, public safety, and the public interest.²⁵

Sandoval, here, focuses on the notion of transparency but gives it a robust and active linkage to the idea of encouraging greater (and more meaningful) access to the internet. She recommends, first, exploring whether consumers are properly informed about ISP practices to enable consumer choice; and second, that the government collect more data on the existence of ISP throttling—and who it affects and how.²⁶ Finally, Sandoval argues, “[c]onsistent with corporate pledges to promote equity and inclusion, ending ISP practices that close the digital schoolhouse, healthcare, and economic opportunity door by throttling users back to the 90s would enable equity, inclusion, public health, and public safety.”²⁷

III. THE IMPACT OF UNDERREGULATION

Sandoval’s observation above brings us to the second principle of *Lex Reformatica*, drawn from a collection of these essays: reformers must recognize the negative impact of technology underregulation on individual civil rights, like privacy, due process, and equality. While many can remain anxious that needed reforms might imperil the freedom that originally defined the frontiers of cyberspace, it bears mentioning, as Gautam Hans has pointed out, that “the internet is already not functioning in so many obvious ways,” pointing out that the current regime of decentralized, uneven regulation has produced troubling consequences.²⁸

While *Lex Informatica* was written right after the dawn of the internet, we might view the papers by Tiffany Li, Gautam Hans, and Catherine Sandoval as a collection of studies that astutely demonstrate the negative impacts of a primary commitment to marketplace control. In a variety of circumstances, our commentators have shown us that the rise of technological norms, when unencumbered by close regulation of the marketplace, can flourish at the cost of equality, privacy, and due process.

Consider, for example, privacy law as an example of this trajectory. Reidenberg was known first and foremost as a privacy scholar. Had Reidenberg been alive today to see the degree of federal inaction in protecting

25. *Id.* at 132.

26. *Id.* at 133.

27. *Id.* at 134.

28. G.S. Hans, *Revisiting Roommates.com*, 36 BERKELEY TECH. L.J. 1227, 1250 (2021).

privacy as a civil right, he would be fairly disappointed, particularly in light of his substantial European work affirming the link between privacy protection and regulation. As Tiffany Li has argued in her contribution to this Issue, privacy is a civil right and yet carries a kind of internal unevenness: as a body of principles, privacy law has failed to account for its *own* inequality in the sense that different people enjoy different levels of protection, both in type and intensity.²⁹ The more that we move into online spaces, Li points out, the more we are connected and the more opportunities there are for our civil rights to be violated in non-traditional ways.³⁰

This is true, not just regarding privacy, but also regarding anti-discrimination as well. Take, for example, the case of *National Fair Housing Alliance v. Facebook*, discussed by Li.³¹ In that case, Facebook permitted advertisers to selectively exclude racial segments of the population from viewing housing ads.³² As Li points out, Facebook was able to discriminate against certain groups by relying on targeted advertising. Yet since targeted advertising relies intrinsically on the practice of data collection, Facebook's action was also a type of downstream harm affecting privacy because it only arises by virtue of the invasive data policies that enable racial categorization in the first place.³³

Li's deft weaving of privacy and antidiscrimination concerns highlight a crucial distinction between privacy protections that function as civil liberties versus privacy protections that function as civil rights.³⁴ As she argues, "unequal access to privacy is a civil rights problem."³⁵ If we only envision privacy as a civil liberty, we miss the interplay between privacy and equality, doing a disservice to both realms.³⁶ In order to protect both, Li argues that we must reconceive of privacy as integral to both due process and equal protection:

Laws and practices that promote surveillance, mandate the use of biased algorithmic assessment, and allow for gendered harms related to cyberstalking, should also be considered unconstitutional based on due process and equal protection. Individuals should be able to

29. Tiffany C. Li, *Privacy As/And Civil Rights*, 36 BERKELEY TECH. L.J. 1265, 1269 (2021).

30. *Id.* at 114.

31. *Id.*

32. *Id.* at 115 (citing Complaint, Nat'l Fair Hous. All. v. Facebook, Inc., No. 1:18-cv-02689 (S.D.N.Y. Feb. 6, 2019)).

33. *Id.* at 115.

34. *Id.* at 108.

35. *Id.* at 109.

36. *Id.* at 110.

claim a constitutional right to privacy under the Equal Protection Clause, recognizing that privacy has never been awarded equally to all people across society.³⁷

Li closes with a call for a federal privacy law that would function to help fill in the gaps that sectoral approaches have left behind, arguing that it would better situate privacy as a civil right, and put us on better footing with other privacy-forward nations, like the EU.³⁸

A related kind of interplay, explored by Gautam Hans in his essay, *Revisiting Roommates.com*, details the relationship between the marketplace, speech protections, and inequality. Here, too, we see the human costs of a failure to integrate antidiscrimination protections into our efforts to regulate the internet. As has been widely discussed in technology law literature, Section 230 immunizes websites from liability for publishing unlawful speech that is made by third parties on their owned platform.³⁹ The reasoning behind this safe harbor is relatively straightforward: “[a]t the scale at which the platforms hope to operate,” Hans explains, “the potential for liability would be immense, as would the costs of prescreening content.”⁴⁰

Yet in *Roommates.com*, two organizations filed suit against Roommates.com, contending that the company violated housing discrimination laws by mandating that end users fill out questionnaires that required disclosure of a user’s age, gender, sexual orientation, and familial status, all of which are identity categories protected by federal fair housing laws.⁴¹ In that case, the Ninth Circuit found that since Roommates.com facilitated connections between third parties (mandating that users answer questions that could result in discrimination), it could not be completely absolved from liability.

Yet Hans’s intervention, however, imaginatively forces us to explore how *Roommates.com*, as an entity, would have fared in the absence of the protective sphere of Section 230. This move, in turn, asks us to situate *Roommates.com* alongside the history of housing discrimination and tenants’ rights that characterizes the intersection of civil rights and housing discrimination.⁴² Just as Li’s paper asks us to imagine a broader framing of privacy law (and its attendant inequalities), Hans asks us to broaden our framing of equality principles in order to reimagine antidiscrimination protections among

37. *Id.* at 124.

38. *Id.* at 125.

39. *See* Hans, *supra* note 28, at 103.

40. *Id.*

41. *Id.* at 107.

42. *Id.* at 105.

platforms. As Hans points out, the import of Section 230 essentially enables platforms to escape liability merely because of their online status, as opposed to newspapers, for example, which would have been required to comply with the Fair Housing Act.⁴³ But it's equally revealing, Hans points out, that civil rights concerns were excluded from its list of exceptions to immunity, an exclusion which has the less desirable effect of potentially creating an exception "large enough to potentially swallow" antidiscrimination protections entirely.⁴⁴ Hans concludes by calling for Section 230 reform that expands the role it could play in the civil rights and racial justice movements; supporting, rather than overruling, the goals of the Fair Housing Act.⁴⁵

Again, both papers highlight the aftereffects of decades of federal underregulation, illustrating the negative externalities that affect individuals and their civil rights—regarding both privacy and equality.

IV. THE DESIGN OF REFORM

A third principle in the *Lex Reformatica* landscape involves the concept of design as a stand-in for direct regulation. The notion of design-oriented solutions is a thread that weaves through many of the papers but is most explicitly explored by Hans (among others), who invokes design-oriented solutions as one way to remedy the inequalities that flow from immunity under Section 230. One solution to the quandary he explores involves a powerful refiguring of design choices—Section 230 could be reformed to create a potential opening for liability for design decisions (i.e., the choices a company makes in setting up its system)—such as its design of its drop down menus, pre-screening its content for liability issues, and the like.⁴⁶ In this way, a company would be responsible, *ex ante*, for ensuring that its design choices do not invite discrimination by its users.⁴⁷

Hans' invocation of design reform, in many ways, indirectly echoes many of the same insights that have been associated with the emergent design justice movement, a concept associated with Sasha Costanza-Chock.⁴⁸ A general definition of the movement is "a field of theory and practice that is concerned with how the design of objects and systems influences the distribution of risks,

43. *Id.* at 113–14.

44. *Id.* at 116.

45. *Id.* at 122–25.

46. *Id.* at 122.

47. *Id.* at 123.

48. See Sasha Costanza-Chock, *Design Justice, A.I., and Escape From the Matrix of Domination*, J. DESIGN & SCI. (July 16, 2018), <https://jods.mitpress.mit.edu/pub/costanza-chock>.

harms, and benefits among various groups of people.”⁴⁹ Here, special attention is paid to whether design reproduces (or is reproduced by) matrices of domination.⁵⁰ The concept of design justice is also oriented normatively, in that it works to build solutions that ensure fair and meaningful participation in design decisions and to recognize community based design and practice.⁵¹ Overall the concept of design justice seeks to provide a “more equitable distribution of design’s benefits and burdens,” along with meaningful participation in design choices, and recognizes the value of community based practices.⁵²

Notably, Hans’ treatment of design reform (and the association it indirectly draws with the notion of design justice) highlights yet another core aspect of *Lex Informatica*: the idea that “law is not the only source of rules or rulemaking.”⁵³ As Margot Kaminiski explains, drawing from Reidenberg: “[t]echnological architecture is its own distinct regulatory force. This insight has serious implications for the law. It means.... That technology isn’t understood to be value-neutral, authoritative, or inevitable. It reflects choices. It’s political.”⁵⁴

As she argues, Reidenberg was “the first to say that architecture mattered.”⁵⁵ But the interaction, she points out, between law and technology does not have one singular formation; it can take on a variety of different formations. Here, law can structure the development of certain technologies, be tied closely to policy goals and values, or make salient particular aspects of doctrine.⁵⁶ Law, in this sense, operates to construct technology into its own systems of meaning and value.⁵⁷ “The internet isn’t a no-lawyer’s land; it’s

49. Sasha Costanza-Chock, *Design Justice: Towards an Intersectional Feminist Framework for Design Theory and Practice*, in 2 DESIGN AS A CATALYST FOR CHANGE—DRS INT’L CONFERENCE 2018 529, 529 (Cristiano Storni, Keelin Leahy, Muireann McMahon, Peter Lloyd & Erik Bohemia eds., 2018); see also Sasha Costanza-Chock, DESIGN JUSTICE: COMMUNITY-LED PRACTICES TO BUILD THE WORLDS WE NEED (2020).

50. *Id.* at 533.

51. The concept of design justice grew out of a summit of designers, artists, technologists, and community organizers in a 2015 meeting of the Allied Media Conference. *Id.* at 529.

52. *Id.* at 533.

53. Margot E. Kaminiski, *Technological “Disruption” of the Law’s Imagined Scene: Some Lessons from Lex Informatica*, 36 BERKELEY TECH. L.J. 883, 886–887 (2021).

54. *Id.*

55. *Id.*

56. *Id.* at 110–12.

57. *Id.* at 112.

populated by people who need stability, norms, rules, and consequences. People *need* the law and *grow* the law; it isn't imposed upon them."⁵⁸

While Reidenberg recognized that law's reach would be inevitable over the internet's inherent messiness, Kaminski subtly exhorts us to return to the messiness of technology in order to excavate how technology changes the "imagined regulatory scene," or the law and policy conversations that are imagined to take place.⁵⁹ For Kaminski, the rise of technology and its architecture can disrupt these imagined scenes, forcing us to examine and reconstruct the nature and justification of regulation itself—as she writes, "not just the 'how' and 'what' of law, but also the 'why.'"⁶⁰ As she points out, the very meaning of technology takes on a certain particularity in tandem with "what one thinks the law is or should be"; in other words, the law constructs technology, just as it constructs other categories of meaning.⁶¹ Here, we should interrogate the balancing of these values and categories, asking constantly whether or not we have struck the correct equilibrium and exposing the particular values that we are acting upon.⁶²

V. CRITICAL EXAMINATION OF THE INTERPLAY OF PUBLIC AND PRIVATE REGULATION

Kaminski's description of the law's "imagined scene" brings us nicely to the fourth aspect of *Lex Reformatica*: the importance of critically examining the interplay between public and private forms of regulation. Whereas Reidenberg's work was characterized by a clear dividing line between public and private forms of architectural regulation, our more modern era has revealed that there are substantial slippages and boundaries between these forms. Private forms of regulation permeate the internet, as content moderation and other strategies of compliance have flourished in the last thirty years. At the same time, just as Reidenberg implicitly predicted, these private forms of regulation have emerged, just as public forms of regulation have largely faded into the background of technology, until just recently.

This is not entirely by accident. As Julie Cohen describes in this Issue—in comparing Lessig's *Code* to *Lex Informatica*, one is struck by the divergence of

58. *Id.* at 106.

59. *Id.* at 104.

60. *Id.* at 131.

61. *Id.* at 112.

62. *Id.* at 131.

their approaches.⁶³ As she explains, because of Reidenberg's exposure to North American and European ways of approaching law and regulation, he focused his gaze on the mechanics of integrating regulatory authority into the development of technologies.⁶⁴ In contrast, *Code*, being a product of Lessig's training at the University of Chicago, focused much more on the role of social norms and markets in influencing the design of technologies.⁶⁵ She writes:

And yet "Lex Informatica," but not *Code*, surfaced the complex *interplay* between regulatory forces. "Lex Informatica" framed new digital formations as situated opportunities for interventions by policymakers and other interested actors—an approach broadly compatible with decades of accumulated, interdisciplinary learning on emergent sociotechnical processes—whereas *Code* described an elemental regulatory struggle that unfolded as a contest over *terra nullius* and that resonated with the reigning neoliberal ethos of the era.⁶⁶

As Cohen eloquently observes, both *Code* and Lex Informatica were ill-equipped to handle the unexpected developments that came with the evolution of the relationship between law and technology.⁶⁷ For example, how does one define "compliance" in an age of algorithmic processes?⁶⁸ What legal obligations do platforms face in addressing the activities of their end users?⁶⁹ How do we ensure public accountability over compliance operations, Cohen asks.⁷⁰ How do we ensure citizens' privacy and dignity in a world of data-driven surveillance?⁷¹

This brings us to a fourth element in the architecture of an era of Lex Reformatica: a critical focus on the interplay of public and private forms of regulation (and self-regulation). In today's contexts, we are grappling with the rise of private compliance systems, systems that are designed to both satisfy legal parameters at the same time that they can often enable circumvention of civil rights principles. In Cohen's masterful exploration of organizational transformation, she argues that "networked information technologies are not

63. Julie E. Cohen, *From Lex Informatica to the Control Revolution*, 36 BERKELEY TECH. L.J. 1017, 1019 (2021).

64. *Id.*

65. *Id.*

66. *Id.*

67. *Id.*

68. *Id.*

69. *Id.*

70. *Id.* at 104.

71. *Id.*

simply new modes of knowledge production to be governed, but also powerful catalysts for organizational restructuring that change the enterprise of governance . . . from the inside out.”⁷²

In the context of platform speech, for example, Cohen brilliantly elucidates how a reliance on probabilistic and engagement metrics in online communities drives users toward extremist speech, essentially foreclosing the efficacy of content or speaker interventions to disrupt their mechanisms.⁷³ Among copyright conflicts and otherwise, speech and content platforms have emerged and evolved, elevating generative speech and content over more rigorous forms of gatekeeping.⁷⁴ And under these circumstances, Cohen argues, civil rights advocates are stuck, unsuccessfully trying to bargain with tech giants who wrap many of their activities within a thick shroud of secrecy, foreclosing mutually acceptable forms of compromise.⁷⁵ Not surprisingly, amidst this climate, a variety of kinds of privatized governance emerge, producing “vast new compliance industries dedicated to the pursuit, perfection, and legitimation of self-governance.”⁷⁶ While Cohen’s description might lead us to consider the failures and limitations of these efforts, she ends on a note that encapsulates Reidenberg’s persistent optimism: center innovation in the law; consult private industry but remain skeptical, that is, avoid equating self-interested positioning with the importance of human flourishing; and finally, “remember that law is a means to an end.”⁷⁷

Whereas Cohen offers us a broad, abstract view of the relationship between private and public forms of regulation, Deirdre Mulligan and Ken Bamberger take a narrower, functional approach in their study of content moderation. Here, the authors bravely—and deeply—engage with the porous and shifting strands of public and private content moderation, arguing that it involves distributed forms of public and private oversight that are, in turn, mostly delegated to a diverse range of actors. Drawing on a range of case studies—the DMCA, the General Data Protection Regulation (GDPR), the governance of online material tied to child sexual abuse, Section 230, and the right to be forgotten—revealing how content moderation tools delegate and constrain decision-making by private actors through deploying a typology of

72. *Id.* at 111.

73. *Id.* at 113–14.

74. *Id.* at 117–18.

75. *Id.* at 120–22.

76. *Id.* at 125.

77. *Id.* at 134.

subfunctions.⁷⁸ By looking to these actions, which they describe as *defining*, *identifying*, *locating*, and *moderating*, Mulligan and Bamberger then use these subfunctions to pose a deeper interrogation of these forms of private governance in light of the public values of accountability.⁷⁹

Drawing in part on a previous collaboration with Helen Nissenbaum and the work of the New Governance school of thought, which focuses on the decentering of public forms of governance by private actors,⁸⁰ the authors argue that in order to truly understand the values promulgated by content moderation systems, we need to identify, separate, and examine the various subfunctions that operate in content moderation and the various hand-offs that take place between human and technical means.⁸¹ In applying this insight to a broad array of mechanisms, all of which focus on content moderation, the authors beautifully lay out the various critiques associated with each form and the ethical and political questions that arise.⁸² To ameliorate some of the disadvantages of these subfunctions, particularly regarding various degrees of definitional competency, the authors profitably argue in favor of a kind of transparency that takes into account the need to disclose the definitions and decisional criteria used in content moderation, and to foster the participation of other stakeholders.⁸³ Other considerations that they describe involve developing more competencies through focusing on the cultural contexts that surround content moderators and their fit with the content that is being regulated.⁸⁴

As Mulligan and Bamberger demonstrate, a close eye to the intricacies of the functions—and subfunctions—of content moderation can yield important insights into the risks and benefits behind private forms of content moderation. But it is important to note, as Hanna Bloch-Wehba points out in her essay, the porosity of the relationship between public and private, perhaps unintentionally, creates a situation, the extent to which forces us to confront the effect of content moderation on law enforcement and vice versa. As she argues, “the purportedly private rules of content moderation are created and

78. Deirdre K. Mulligan & Kenneth A. Bamberger, *From Form To Function in Content Moderation*, 36 BERKELEY TECH. L.J. 1091, 1110 (2021).

79. *Id.* at 106.

80. *Id.* at 115.

81. *Id.* at 115–16 (citing Deirdre K. Mulligan & Helen Nissenbaum, *The Concept of Handoff As a Model for Ethical Analysis and Design*, in OXFORD HANDBOOK ETHICS & AI 234 (Markus D. Dubber, Frank Pasquale & Sunit Das eds., 2020).

82. *Id.* at 107.

83. *Id.* at 153.

84. *Id.* at 156.

operate within a political context in which law enforcement acts as a particularly powerful stakeholder.”⁸⁵

This produces a dialectic between law enforcement and social media content that constructs both trajectories in turn: content moderation, i.e., speech surveillance on platforms, shapes the scale and design of law enforcement and vice versa. “Just as law enforcement seeks expanded influence over platforms’ private decision-making, the processes and technical affordances of content governance also affect and shape law enforcement investigations in more mundane contexts.”⁸⁶ Here, while law enforcement aims to influence content moderation on platforms, the substantive, procedural, and technical rules that govern platforms shape law enforcement itself.⁸⁷ She studies the complex relationship between government pressure to moderate content and private platforms’ responses, noting how government influence can shape private content moderation systems.⁸⁸ In turn, the private decision-making (including standard setting) activities of firms can also shape the behavior of law enforcement in seeking data and user information, even arguably making it more difficult to procure content involving illegal activity.⁸⁹

In making these observations, Bloch-Wehba offers another crucial variable for consideration in the discussion of private and public forms of regulation: the self-interested nature of law enforcement. As she brilliantly elucidates, the technological modalities that govern online content can also feed the interests of law enforcement, creating “new types and sources of information relevant to new kinds of investigations.”⁹⁰ In contexts as varied as terrorist content and sex work, private platforms essentially operate as proxy censors; even though platforms enjoy some modicum of immunity, they are still indirectly pushed by law enforcement to behave more aggressively in filtering unlawful content.⁹¹ As she explains:

In the context of terrorist imagery, platforms are required to report certain kinds of terroristic threats to European authorities and likewise required to preserve a broader range of information for future law enforcement use. The result is that the monitoring technology used to detect lawbreaking itself lies at the heart of

85. Hannah Bloch-Wehba, *Content Moderation as Surveillance*, 36 BERKELEY TECH. L.J. 1297, 1300 (2021).

86. *Id.*

87. *Id.* at 106.

88. *Id.* at 110–11.

89. *Id.* at 105.

90. *Id.* at 106.

91. *Id.* at 117.

investigations and prosecutions, yielding increasing entanglements between law enforcement and platform governance.⁹²

The creeping shadow of law enforcement over these privatized forms of governance, Bloch-Wehba points out, forces us to confront how the implicit, indirect collusion between the two dismantles the prevalent assumption that private platforms wholly engage in self-governance and are accountable to no one but their shareholders.⁹³ The “extensive alignment” of platforms and law enforcement lacks the kind of accountability that our legal system is premised upon; twenty-three years after *Lex Informatica*, Bloch-Wehba notes, “U.S. law has made little progress in ensuring that *lex informatica* is as democratically legitimate or accountable as its regulatory equivalents.”⁹⁴ Hence the need for a more critical, and searching, inquiry into future possibilities for its governance.

VI. EX ANTE REGULATION VS. EX POST REMEDIES

Much of the critical points made by our authors in this Issue have pointed their gaze toward the interplay between private and public, pointing out (as Bloch-Wehba has done) how various incentives and interests can often complement one another, often at the cost of accountability to the public. In order to address this issue, Reidenberg himself advocated “a shift in the focus of government action away from direct regulation and toward indirect influence” by regulating behavior and standards before even the consequences of technology surface.⁹⁵

This important insight about the value of proactive intervention, rather than after-the-fact forms of correction, operates at the heart of many of the essays in this Issue, but it appears most directly in Ifeoma Ajunwa’s powerful study of the increasing use of automated video interviewing and the issues that it raises for employment discrimination.⁹⁶ She points out that these technologies are often touted as anti-bias interventions but can paradoxically run the risk of not just replicating the bias they are meant to evade, but also amplify it.⁹⁷

Ajunwa’s insightful case study threads a number of themes that I’ve discussed regarding the collection of essays, but her work demonstrates yet an

92. *Id.* at 133–34.

93. *Id.* at 141.

94. *Id.* at 141, 144.

95. Ifeoma Ajunwa, *Automated Video Interviewing as the New Phrenology*, 36 BERKELEY TECH. L.J. 1173, 1218 (2021) (citing Reidenberg, *supra* note 1, at 586).

96. *See generally id.*

97. *Id.* at 103.

additional principle: the value of ex ante legal regulations as opposed to ex post remedies.⁹⁸ As Ajunwa has noted in her application of Reidenberg, one critical benefit of Lex Informatica is that it relies on ex ante measures of execution.⁹⁹ Instead of remedying harm that has already occurred, Lex Informatica enables automated monitoring and enforcement even before the violation has even taken place.¹⁰⁰ Here, she draws directly on Reidenberg to show how thoughtful, proactive interventions at the design stage can radically deter discrimination at a later stage.¹⁰¹

Ajunwa's article, here, is a masterful example of the risks and benefits of the AI-driven era we inherit today. Consider the case study that she relates, so vividly: (1) a new technology is invented (here, automated video interviewing) that seems to increase efficiency but that actually perpetuates bias; and (2) existing legal principles make it difficult to establish a case of illegal discrimination under existing protections (Title VII, the ADA, and various informational privacy entitlements). In such situations, Ajunwa argues that a Lex Informatica framework might push a more proactive intervention at a much earlier stage: "Whereas traditional law would require candidates to know a violation of their rights occurred in order to seek protection—a serious problem given the opaque nature of algorithmic decision-making—technological solutions under a Lex Informatica framework provide some assurance that such violations will not occur in the first place."¹⁰²

Again, by enlisting design principles at an early stage, Ajunwa invokes Lex Informatica, arguing that the Uniform Guidelines on Employee Selection Procedures should be employed in the design of such systems, including the collection, validation, and use of particular content.¹⁰³ She even argues that Lex Informatica might provide the basis for a property entitlement in the subject's informational privacy interests, again showing us how an ex ante design orientation might foreclose the risk of bias and discrimination at a later stage.¹⁰⁴

Ajunwa's insightful case study offers us a lesson in employing Lex Informatica in the very fraught area of recruitment and employment. While she shows us how a regulatory framework could be employed at the front end to govern and confront the potential bias that surfaces from one form of a

98. *Id.* at 104.

99. *Id.* at 146–47.

100. *Id.*

101. *Id.* at 104.

102. *Id.* at 146–47.

103. *Id.* at 147–48.

104. *Id.* at 151–53.

new technology, in our final paper in this Issue, Karni Chagal-Feferkorn and Niva Elkin-Koren focus on a similar pattern emerging from a much broader array of technologies that they collectively refer to as Lex AI.¹⁰⁵ While there has been much ink spilled describing the domain of AI, the authors treatment reframes our gaze towards addressing not the issue of solving the problems of AI but rather *how* AI governs human behavior.¹⁰⁶

Refreshingly, they argue, Lex Informatica has given way to another system of private ordering through technology, a system that includes personalized recommendations and other forms of data-driven decision-making.¹⁰⁷ Yet here, the authors stop short of describing Lex AI as just another form of private ordering; instead, they argue that it comprises an enabler of collective action.¹⁰⁸ The paper, like those mentioned in the Section above, carefully revisits the public/private distinction, especially in law and economics literature, which is often deployed to justify the appropriate scope of regulatory intervention. Yet here, the authors perceptively ask how to approach governance in such circumstances, particularly given its infrastructure which might seem (at first glance) to aggregate the will of connected individuals, but in fact (the authors point out) functions as a robust, unaccountable, mechanism that actively constructs, rather than collects, the will of individuals.

Unlike many of the other authors in this Issue, who undertake a notably critical gaze towards AI and its aftereffects, Chagal-Feferkorn and Elkin-Koren choose to instead consider Lex AI as a kind of sui generis “unicorn” type of governance: one that invites closer scrutiny because it lacks some of the typical advantages of private ordering, while also typifying some of the problems of collective action.¹⁰⁹ The authors draw an insightful comparison between Lex Informatica and decision-making by AI, arguing that the latter could be viewed as a system of governance because of the way that it generates norms and affects the behavior of users but that it can also be deployed by private or public entities.¹¹⁰ Here, Lex AI also enables a greater degree of personalization in a variety of different areas, including decision-making in both the private and public spheres.¹¹¹

105. Karni A. Chagal-Feferkorn & Niva Elkin-Koren, *Lex AI: Revisiting Private ordering by Design*, 36 BERKELEY TECH. L.J. 915, 919 (2021).

106. *Id.* at 107.

107. *Id.*

108. *Id.* at 108.

109. *Id.* at 109.

110. *Id.* at 117–18.

111. *Id.* at 121–22.

Perhaps the biggest payoff from their thoughtful account lies in their unwillingness to analogize Lex AI to a complete form of either public or private ordering. As they argue, while Lex AI “can be easily mistaken for a private ordering form of governance, [Lex AI] is in fact closer in nature to public ordering.” But at the same time, the analogy is incomplete; as they show, Lex AI also lacks some of the key characteristics at the basis of public ordering, as well. AI drives its predictions from a centralized process of decision-making, leading the authors to conclude that it may resemble a distinctive type of collective action mediated by algorithms, rather than by self-governance.¹¹² As a result, Lex AI cannot qualify as strictly top-down governance, since it is formed by data-driven predictions, which are dynamic, distributed, and far less predictable than traditional modes of “command-and-control governance.”¹¹³

Critically, the authors are careful to note that Lex AI is permeated by a particular type of power asymmetry. At the same time that it offers efficient ways to manage and analyze data, Lex AI cannot always reflect a user’s personal choice.¹¹⁴ The authors note, “Lex AI does not provide a reliable signaling of people’s preferences and choices. The way preferences are inferred and the recursive process by which Lex AI shapes norms and behaviors may result in predictions that fail to reflect individuals’ true preferences and may generate inefficiencies.”¹¹⁵

In the end, the authors argue that the advent of Lex AI is wholly different than traditional forms of governance—it is not exactly centralized, nor is it totally distributed.¹¹⁶ And this insight—one might characterize it as a refusal to analogize—offers a host of possibilities for a new way to approach governance. The authors close by suggesting that public policy “treat Lex AI as an ecosystem [shaped by] various types of (sometimes unidentified) entities,” each of which carries the capacity to shape the adaptive learning of the centralized system, thereby affecting the decisions and predictions that it produces.¹¹⁷ By including a broader array of stakeholders and by questioning the validity of data-related and design choices, Lex AI can better perfect the regulatory and legal measures that it may generate.¹¹⁸

112. *Id.*

113. *Id.*

114. *Id.*

115. *Id.*

116. *Id.*

117. *Id.* at 145–46.

118. *Id.*

Of course, the central question that threads through their article, along with each of those I've discussed in this Foreword, involves the question of governance. Because of its data-driven, dynamic nature, the authors point out that *ex ante* scrutiny of its systems may be simply impossible to perform.¹¹⁹ Because its adaptive functions are driven by data, they are inherently dynamic; the norms that embed the system are opaque and nondiscursive, short-circuiting the opportunity for an interrogation of its values. Even more than its opacity, or in part because of it, Lex AI fails to facilitate a discursive engagement into how it functionally corresponds to social norms and values and thus fails to enable a deeper public deliberation of its utility.¹²⁰

In the end, as they gently warn, our present inability to grapple with a system of governance of Lex AI may result in at least a partial undoing of—or at least a challenge to—democracy. But as they point out, the future does not necessarily have to be this way. Indeed, as they describe at the end of the essay, perhaps AI may result in a redoing (as opposed to an undoing) of democracy, if we can reconfigure our thinking:

Treating Lex AI as a data ecosystem rather than a rule-based design would assist in focusing attention on the interaction between the different actors when considering legal tools to mitigate potential harms the system generates.¹²¹

For these authors, the mitigation of AI-related harms requires us to train our lens towards a wider gaze; an inquiry that considers more than code alone, and also reckons with the potential limitations inherent in the sources of data, as well as its recursive and dynamic effects on decision-making and governance.¹²² As they warn, under this framing of an ecosystem, “design and deployment decisions pertaining to the system might be affected by different stakeholders” but should not be interpreted to function as a shield from liability for those choices.¹²³ Indeed, it is the opening of these possibilities—design choices, deployment choices, the interaction between stakeholders and the potential for liability for harm—that carries the greatest *ex ante* potential for improving regulatory and legal measures in the future.¹²⁴

119. *Id.* at 145.

120. *Id.*

121. *Id.*

122. *Id.* at 147.

123. *Id.* at 148.

124. *Id.*

VII. CONCLUSION

These articles, as insightful and varied as they are, collectively represent a necessary conversation between the wisdom of prior generations of technology scholars like Joel Reidenberg, who focused on the possibilities of fairness through customs, norms, and the design of technology; and a newer generation of technology scholars who rightfully draw our attention to the aftereffects of an absence of regulation on vulnerable groups. Collectively, these papers represent an unfolding conversation about technology, norms, design, and regulation—a world of *Lex Reformatica* that Reidenberg himself would have deemed full of possibilities for meaningful integration in the future.

TECHNOLOGICAL “DISRUPTION” OF THE LAW’S IMAGINED SCENE: SOME LESSONS FROM *LEX INFORMATICA*

Margot E. Kaminski[†]

ABSTRACT:

Joel Reidenberg in his 1998 Article *Lex Informatica* observed that technology can be a distinct regulatory force in its own right and claimed that law would arise in response to human needs. Today, law and technology scholarship continues to ask: does technology ever disrupt the law? This Article articulates one particular kind of “legal disruption”: how technology (or really, the social use of technology) can alter the imagined setting around which policy conversations take place—what Jack Balkin and Reva Siegal call the “imagined regulatory scene.” Sociotechnical change can alter the imagined regulatory scene’s architecture, upsetting a policy balance and undermining a particular regulation or regime’s goals. That is, sociotechnical change sometimes disturbs the imagined paradigmatic scenario not by departing from it entirely but by constraining, enabling, or mediating actors’ behavior that we want the law to constrain or protect. This Article identifies and traces this now common move in recent law and technology literature, drawing on Reidenberg’s influential and prescient work.

TABLE OF CONTENTS

I.	INTRODUCTION	884
II.	UNIFYING PRINCIPLES FOR LAW AND TECHNOLOGY: THE INSIGHTS OF <i>LEX INFORMATICA</i>	886
III.	“LEGAL DISRUPTION”	892
IV.	“LEGAL DISRUPTION” AND THE IMAGINED REGULATORY SCENE	895
	A. IMAGINED REGULATORY SCENES	895
	B. ... AND ARCHITECTURAL DISRUPTION	897
	C. IMPLICATIONS	903
	D. THE MOVE IN THE LITERATURE	905

DOI: <https://doi.org/10.15779/Z38JW86N97>

© 2021 Margot E. Kaminski.

[†] Associate Professor of Law, Colorado Law. Director of the Privacy Initiative at Silicon Flatirons. Many thanks to Ryan Calo, Julie Cohen, Rebecca Crootof, and Meg Jones for helpful comments. Immeasurable thanks to the late Joel Reidenberg for his kindness, support, scholarship, and leadership in the field. Mistakes are my own.

1.	<i>Architectural Changes as Constraints</i>	905
2.	<i>Architectural Changes as Affordances</i>	908
3.	<i>Architectural Changes as Mediation or Channeling</i>	911
V.	CONCLUSION	913

I. INTRODUCTION

In 1996, as the keynote at an early cyberlaw conference, Judge Frank Easterbrook famously characterized internet law as “the law of the horse.”¹ Throwing down the gauntlet for generations of technology lawyers and professors to come, Judge Easterbrook explained that what was then known as cyberlaw was, like most studies of objects or actors affected by the law, a collection of questions from disparate areas of legal practice. “Any effort to collect these strands” into a law school class—or, by implication, a discipline—“is doomed to be shallow and to miss unifying principles.”²

He was wrong.

In 1998, Joel Reidenberg provided a set of unifying principles and practices for technology law.³ Many of those principles still hold true today. Although its borders may have changed and its corpus(es) of substantive law expanded, technology law as a discipline is very much alive and thriving.

This Article begins by charting the core lessons of *Lex Informatica*: the unifying principles of technology law. As Reidenberg observed in 1998, technology can itself be a distinct regulatory force, crafted by extra-legal players in extra-legal institutions. Designing law for technology requires understanding and engaging with extra-legal forces, players, and institutions. Writing at a time when frontier metaphors infused a lot of the early scholarship,⁴ Reidenberg claimed instead that law would be not evadable but inevitable, arising as the natural consequence of human social practices and needs.⁵

1. Frank H. Easterbrook, *Cyberspace and the Law of the Horse*, 1996 U. CHI. LEGAL F. 207, 208 (1996).

2. *Id.* at 207.

3. Joel R. Reidenberg, *Lex Informatica: The Formulation of Information Policy Rules Through Technology*, 76 TEX. L. REV. 553 (1997).

4. See, e.g., John Perry Barlow, *A Declaration of the Independence of Cyberspace*, (Feb. 8, 1996), <https://www.eff.org/cyberspace-independence>; David R. Johnson & David Post, *Law and Borders—The Rise of Law in Cyberspace*, 48 STAN. L. REV. 1367 (1996).

5. Reidenberg, *supra* note 3, at 553–54 (comparing the necessary evolution of technology law to the evolution of *Lex Mercatoria* and calling for “ground rules” to “offer

This Article next turns to recent dialogue from the field of law and technology, which is now decidedly less existential in nature. Rather than trying to justify its own existence, recent law-and-technology scholarship focuses on identifying what makes a particular question interesting, versus the practice of law as usual.⁶ In other words, it asks: does technology ever disrupt the law?⁷ And if so, when and how?

stability and predictability so that participants have enough confidence for their communities to thrive, just as settled trading rules gave confidence and vitality to merchant communities”).

6. See *infra* note 7.

7. Even framing the question this way is controversial as disruption implies a one-way arrow of influence of law on technology. Also, when we refer to “technology,” we rarely mean an object and almost always mean social adoption and uses of technology. The recent and ongoing debate about technological exceptionalism and technological determinism in the field asks whether a particular technology has special qualities that make it disruptive to the law or whether there are other ways to characterize “disruption.” See, e.g., Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CALIF. L. REV. 513, 515 (2015) (“Robotics is shaping up to be the next transformative technology of our time. And robotics has a different set of essential qualities than the Internet. . . . The essential qualities of robotics will drive a distinct conversation [about the law.]”); Jack M. Balkin, *The Path of Robotics Law*, 6 CALIF. L. REV. CIR. 45, 45 (2015) (“I do not think it is helpful to speak in terms of ‘essential qualities’ of a new technology that we can then apply to law. On the contrary, we should try not to think about characteristics of technology as if these features were independent of how people use technology in their lives and in their social relations with others. Because the use of technology in social life evolves, and because people continually find new ways to employ technology for good or for ill, it may be unhelpful to freeze certain features of use at a particular moment and label them ‘essential.’”); Meg Leta Jones, *Does Technology Drive Law? The Dilemma of Technological Exceptionalism in Cyberlaw*, 2018 U. ILL. J. L. TECH & POL’Y 249, 253 (“I argue that technology does not drive law either. Technology is not the locus of legal agency. When testing the theory of technological exceptionalism, no technology has even been exceptional. We must figure out a new way to answer the question, ‘are driverless cars new?’ Because, [sic] technological exceptionalism is not up to the task. Instead of analyzing whether technologies are or will be exceptional and in addition to analyzing how the law can and should respond to exceptional or conservative technological advances, this Article argues that cyberlaw research should consider the way in which technologies, practices, and social arrangements are constructed within certain legal contexts: the legal construction of technology.”); Margot E. Kaminski, *Legal Disruption: How Technology Disrupts the Law* (Mar. 17, 2017) (unpublished manuscript) (on file with author) [hereinafter *Legal Disruption*]; Margot E. Kaminski, *Authorship, Disrupted: AI Authors in Copyright and First Amendment Law*, 51 U.C. DAVIS L. REV. 589, 590–91 (2017) [hereinafter *Authorship, Disrupted*] (“To the extent new technology (or really, the social practice of a new technology) disrupts the law, it does so because of how it encounters existing features of the law, both doctrinal and theoretical. The law, in constructing—that is, building the meaning of—new technological developments and their social uses, takes a central part in its own disruption. Conceiving of technology as some outside force that acts upon the law can lead to a technology-centric approach in which one tries to identify what features of a particular technology are legally disruptive. This kind of disruption narrative gets it wrong. A particular feature of a particular technology disrupts the law only because the law has been structured—doctrinally and theoretically—in a way that makes that feature relevant. The disruptive effects (if any) of a technology become manifest when they encounter, interface

This Article is part of a larger project, one piece in a puzzle that will probably take decades to assemble.⁸ The Article focuses on one particular version of what I have called “legal disruption”: how technology, or really the social use of technology, can change the law’s “imagined regulatory scene.”⁹ That is, each law or policy conversation takes place around an understood imagined setting, with technology, or the lack thereof, often playing a central role. When the social adoption of technology alters the forces in one of these imagined scenes, it can upset the policy balance and undermine the goals of a particular regulation or regime.

Although scholars and policymakers regularly discuss this form of “disruption,” few have identified it as a particular class of analytical move, identified its prevalence in the literature, or discussed that move’s implications and consequences. This Article draws on the lessons of *Lex Informatica* to provide guidance for identifying and addressing technological “disruption” of the law’s imagined scenes—the paradigmatic cases that judges and regulators use to evaluate and interpret the law.

II. UNIFYING PRINCIPLES FOR LAW AND TECHNOLOGY: THE INSIGHTS OF *LEX INFORMATICA*

Reidenberg’s core insight in *Lex Informatica* is that the law is not the only source of rules or rulemaking.¹⁰ Technological architecture is its own distinct regulatory force.¹¹ This insight has serious implications for the law. It means

with, and are given particular meaning within the law.”); Rebecca Crootof & BJ Ard, *Structuring Technolaw*, 34 HARV. J.L. & TECH. 348, 348–49 (2021) (“The conventional approach is to tackle these quandaries by identifying something about a technology or its use that is ‘exceptional’ and argue that this distinction necessitates new law or even a new legal regime; or, alternatively, that a lack of exceptional characteristics implies that the technology can be adequately governed by extant rules. But while these focused studies are individually useful, the exceptionalist approach fosters siloed and potentially incomplete analyses, masks the repetitive nature of the underlying questions, and thereby results in the regular reinvention of the regulatory wheel.”).

8. See, e.g., *Legal Disruption*, *supra* note 7; *Authorship, Disrupted*, *supra* note 7.

9. Jack M. Balkin & Reva B. Siegel, *Principles, Practices, and Social Movements*, 154 U. PA. L. REV. 927, 928 (2006) (“[L]egal principles are intelligible and normatively authoritative only insofar as they presuppose a set of background understandings about the paradigmatic cases, practices, and areas of social life to which they properly apply. A principle always comes with an imagined regulatory scene that makes the meaning of the principle coherent to us. When that background understanding is disturbed the principle becomes ‘unstuck’ from its hermeneutic moorings; it no longer seems clear how the principle applies or even whether it should apply.”).

10. Reidenberg, *supra* note 3, at 554 (“Technological capabilities and system design choices impose rules on participants.”).

11. *Id.* at 555 (describing “technological constraints as a distinct source of rules”).

that there are alternate sources and sites of “rulemaking” that impact technology users as much as, and often more so than, the law. It also means that technology isn’t understood to be value-neutral, authoritative, or inevitable. It reflects choices. It’s political.¹²

Take, for example, internet browsers. Unlike a physical book or magazine, browsers are configured to record a user’s web browsing patterns.¹³ This sets a default rule that personal data will be collected—a default that can be overridden only if the technology is designed to allow for it.¹⁴ As we have seen in the twenty-plus years since Reidenberg published his article, information collection and use have had significant policy consequences, from contributing to digital market manipulation to affecting democratic self-governance at its core.¹⁵

Reidenberg was the first to say that architecture mattered.¹⁶ Technology *is* policy. And architectural policy choices are made usually not by lawmakers but by technologists.¹⁷ Policies are baked into technologies, establishing defaults. Sometimes those defaults are immutable, and sometimes they allow for customization or user choice. Rather than being enforced by the courts, this *Lex Informatica* often enforces automatically—with all the benefits and problems “perfect” enforcement entails.¹⁸

12. See Langdon Winner, *Do Artifacts Have Politics?*, 109 DAEDALUS 121 (1980) (famously discussing the ways in which objects can be characterized as embodying political choices).

13. Reidenberg, *supra* note 3, at 571.

14. *Id.* (“[C]ustomizations through reconfigurations are only possible if the architectural standards support the deviations.”).

15. See, e.g., Neil M. Richards, *Intellectual Privacy*, 87 TEX. L. REV. 387 (2008); NEIL M. RICHARDS, *INTELLECTUAL PRIVACY: RETHINKING CIVIL LIBERTIES IN THE DIGITAL AGE* (2015); Julie E. Cohen, *What Privacy is For*, 126 HARV. L. REV. 1904 (2013); Ryan Calo, *Digital Market Manipulation*, 82 GEO. WASH. L. REV. 995 (2014).

16. Reidenberg was followed by Lawrence Lessig in his famous *Code* and his pathetic (regulated) dot theory. Lawrence Lessig, *The Law of the Horse: What Cyberlaw Might Teach*, 113 HARV. L. REV. 501 (1999) [hereinafter *The Law of the Horse*]; Lawrence Lessig, *The New Chicago School*, 27 J. LEGAL STUD. 661, 662 (1998); LAWRENCE LESSIG, *CODE VERSION 2.0* (2006) [hereinafter LESSIG, CODE].

17. Reidenberg, *supra* note 3, at 569.

18. *Id.* at 569, 576 (discussing PICS-based content filtering as “self-executing” law); see also Edward K. Cheng, *Structural Laws and the Puzzle of Regulating Behavior*, 100 NW. U. L. REV. 655, 716 (2006) (“[D]o we really want a structured society? Are the liberty costs too great? Breaking down the liberty arguments, we see that they largely counsel caution. . . . Structure can infringe on privacy and raise the specter of a police state . . . with Type I structures, which force compliance through surveillance and the constant threat of enforcement. . . . Type II structures may raise accountability concerns because they regulate behavior behind the scenes.”); Harry Surden, *Structural Rights in Privacy*, 60 SMU L. REV. 1605, 1616 (2007) (“One could thus imagine an explicit technological rule built into an automobile’s computer system which limited the maximum speed of the vehicle to a particular value—say 100 miles per hour.

Yet Reidenberg was not a tech determinist.¹⁹ He did not subscribe to the view that technology was ungovernable or that the influence of technology on law flowed only one way. Information technology was, to him, resolutely not some mythological ungovernable frontier. Reidenberg didn't just trust the market. He had faith in good lawmaking, good institutions, and good law.

Lex Informatica affirmatively calls for law to engage directly with technological design. Comparing the new rules of information flows to the norms and customs of sea merchants during the Middle Ages, Reidenberg charts a progressive view that repeated practices naturally become customs, which become agreements, which become law.²⁰ That is, there is a certain inevitability to law, per Reidenberg. The internet isn't a no-lawyer's land; it's populated by people who need stability, norms, rules, and consequences. People *need* the law and *grow* the law; it isn't imposed upon them.²¹ This was not a popular or common view in late 1990s cyber scholarship.²²

Reidenberg was also a pioneer in talking about the importance of law in fostering consumer trust. It has since become common to talk about technology and trust.²³ Nobody wants to get into a driverless car if they know

In the presence of such a structural constraint, and assuming the inability to circumvent the limitation, the speeding behavior could thus be prevented ex-ante, rather than incrementally deterred. In this scenario, the structural constraint is not only self-enforcing—the constraint itself has the ability to detect and prevent violations—it is non-violable.”); Christina Mulligan, *Perfect Enforcement of Law: When to Limit and When to Use Technology*, 14 RICH. J. L. & TECH. 1, 3 (2008) (“When considering whether to use technology to enforce law, a decision-maker should make four determinations. First, is the aversion to using the technology an aversion to the method of enforcing the law or a disagreement with the underlying substantive law? Second, will the technology effectively enforce the law? Third, is the use of the technology constitutional? And finally, does the technology trigger any other philosophical concerns?”); James Grimmelman, *Regulation by Software*, 114 YALE L.J. 1719, 1739 (2005) (“Scholarship on regulatory modalities has convincingly demonstrated that it is often good social policy, rational, and efficient for people to act with indifference to legal rules. Law, extended into such realms, disrupts efficient social norms in those cases in which it is applied. Because software can reach so many more transactions than can law, it can disrupt more cases. And because software, unlike law, is immediate, the effects of each disruption are more severe.”).

19. For a discussion of technology determinism, see Jones, *supra* note 7.

20. Reidenberg, *supra* note 3, at 553 (“Custom and practices evolved into a distinct body of law known as the ‘Lex Mercatoria,’ which was independent of local sovereign rules and assured commercial participants of basic fairness in their relationships.”).

21. *Id.* at 554 (“Principles governing the treatment of digital information must offer stability and predictability so that participants have enough confidence for their communities to thrive, just as settled trading rules gave confidence and vitality to merchant communities.”).

22. See, e.g., *supra* note 4.

23. See, e.g., Ian Kerr, *Personal Relationships in the Year 2000: Me and My ISP*, in RELATIONSHIPS OF DEPENDENCE AND INTERDEPENDENCE IN LAW 78, 110–11 (2002); Helen Nissenbaum, *Securing Trust Online: Wisdom or Oxymoron?*, 81 B.U. L. REV. 635 (2001); Ian R. Kerr, *The Legal Relationship Between Online Service Providers and Users*, 35 CAN. BUS. L.J. 419

driverless cars regularly crash. Users need rules and legal duties so that they can trust in the safety and security of the technologies they use. Back in the mid-1990s, however, it was more common in the United States to call for a wait-and-see approach to technologically enabled harms.²⁴ Reidenberg breaks this trend by calling proactively for “common ground rules to create trust and confidence.”²⁵

Reidenberg emphasizes that technology can be both the problem and at least part of the solution. Once we acknowledge and identify a technology’s politics, Reidenberg explains, we can also use them to align practices with norms and the law.²⁶ For example, privacy-enhancing technologies (PETs) such as cryptography can allow users to make privacy-preserving choices.²⁷ Technical standards and filtering software can be used (albeit, often problematically) to screen for copyrighted or illegal content.²⁸ Reidenberg’s observations back in 1998 prefigure what is now a central element of privacy policymaking: a focus on system defaults and “Privacy by Design.”²⁹ They also prefigure a growing literature on automated enforcement in copyright law.³⁰

Other themes of *Lex Informatica* still reverberate, especially the question of what to do about significant divergence in national policies. Reidenberg writes

(2001); DANIEL J. SOLOVE, *THE DIGITAL PERSON* 103 (2004) (proposing that companies which collect and utilize user personal information be treated as legal fiduciaries); Neil Richards & Woodrow Hartzog, *Taking Trust Seriously in Privacy Law*, 19 STAN. TECH. L. REV. 431 (2016); Jack M. Balkin, *Lecture, Information Fiduciaries and the First Amendment*, 49 U.C. DAVIS L. REV. 1183, 1221 (2016); ARI EZRA WALDMAN, *PRIVACY AS TRUST* 61–76 (2018); Jack M. Balkin, *The Fiduciary Model of Privacy*, 134 HARV. L. REV. FOR. 11 (2020). For a critique of the fiduciary model, see Lina M. Khan & David E. Pozen, *A Skeptical View of Information Fiduciaries*, 133 HARV. L. REV. 497 (2019).

24. *See supra* note 4.

25. Reidenberg, *supra* note 3, at 554.

26. He calls these “policy technologies.” *Id.* at 569, 575.

27. *See id.* at 574 (stating PETs can “facilitate the customized management of information rights in the face of existing technological default rules”).

28. *Id.* at 575–76 (describing the PICS technical standard); Dan L. Burk, *Algorithmic Fair Use*, 86 U. CHI L. REV. 283, 302 (2019) (discussing the problems with relying on filtration software that often incorrectly flags “fair use” of copyrighted works as illegal).

29. *See* Ann Cavoukian, *Privacy by Design: The 7 Foundational Principles* (“Privacy by Design is a concept I developed back in the 90’s . . .”); WOODROW HARTZOG, *PRIVACY’S BLUEPRINT: THE BATTLE TO CONTROL THE DESIGN OF NEW TECHNOLOGIES* (2018); Ari Ezra Waldman, *Privacy’s Law of Design*, 9 U.C. IRVINE L. REV. 1239 (2019); Regulation (EU) 2016/679, of the European Parliament and the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1 at art. 25 [hereinafter GDPR].

30. *See* Maayan Perel & Niva Elkin-Koren, *Accountability in Algorithmic Copyright Enforcement*, 19 STAN. TECH. L. REV. 473 (2016); Matthew Sag, *Internet Safe Harbors and the Transformation of Copyright Law*, 93 NOTRE DAME L. REV. 499 (2017).

that because of the global nature of information flows, “transnational human interactions . . . raise profound conflicts for national and international law,” particularly in the realm of privacy law.³¹ Although in Europe, “comprehensive legal rights exist and government enforcement plays an important role” in data protection regulation, in the United States, “legal rights are limited.”³²

Reidenberg went on to write several seminal works on this transatlantic conflict over privacy law, more than a decade before the European Court of Justice (CJEU) twice invalidated the U.S.-EU transatlantic data transfer agreements (the Safe Harbor and its replacement, the Privacy Shield).³³ Today, transatlantic policy conflicts, both substantive and jurisdictional, remain front and center in information policy, particularly privacy policy.³⁴ We are still debating what to do about divergent norms, rules, laws, and design when information technology flattens the world.³⁵ For example, the CJEU’s 2014 “right to be forgotten” opinion, *Google Spain*, triggered a flurry of recent scholarship assessing fundamental differences between the U.S. and EU approaches to striking a balance between privacy and speech.³⁶

Reidenberg also discusses the so-called “pacing problem,” which remains a central question for law and technology. The question of how to coevolve law with technology or what to do when “technological developments outpace

31. See Reidenberg, *supra* note 3, at 556.

32. *Id.* at 561.

33. Joel R. Reidenberg, *Resolving Conflicting International Data Privacy Rules in Cyberspace*, 52 STAN. L. REV. 5 (2000); Joel R. Reidenberg, *E-Commerce and Trans-Atlantic Privacy*, 38 HOUS. L. REV. 717 (2001).

34. See Paul M. Schwartz & Karl-Nikolaus Peifer, *Transatlantic Data Privacy Law*, 106 GEO. L.J. 115, 117 (2017); Case C-362/14, *Schrems v. Data Prot. Comm’r*, ECLI:EU:C:2015:650, 10–31 (Oct. 6, 2015) [hereinafter *Schrems I*]; Case C-311/18, *Data Prot. Comm’r v. Facebook Ir. Ltd.*, ECLI:EU:C:2020:559 (July 16, 2020) [hereinafter *Schrems II*]. Case C-131/12, *Google Spain SL v. AEPD*, ECLI:EU:C:2014:317, 22 (May 13, 2014) [hereinafter *Google Spain*]; C-507/17, *Google LLC v. CNIL*, 2019 EUR-Lex CELEX No. 62017CJ0507 (Sept. 24, 2019) [hereinafter *CNIL*].

35. See Anu Bradford, *The Brussels Effect*, 107 NW U. L. REV. 1, 22–25 (2012); Anupam Chander, Margot E. Kaminski & William McGeeveran, *Catalyzing Privacy Law*, 105 MINN. L. REV. 1733, 1737–38 (2021).

36. Steven C. Bennett, *The ‘Right to Be Forgotten’: Reconciling EU and US Perspectives*, 30 BERKELEY J. INT’L L. 161 (2012); Robert Post, *Data Privacy and Dignitary Privacy: Google Spain, the Right to be Forgotten, and the Construction of the Public Sphere*, 67 DUKE L.J. 981 (2018); Stefan Kulk & Frederik Z. Borgesius, *Google Spain v. González: Did the Court Forget about Freedom of Expression?*, 5 EURO. J. RISK REG. 389 (2014); see also Hannah Bloch-Wehba, *Global Platform Governance*, 72 SMU L. Rev. 27 (2019); Daphne Keller, *The Right Tools: Europe’s Intermediary Liability Laws and the EU 2016 General Data Protection Regulation*, 33 BERKELEY TECH. L.J. 287 (2018). The CJEU’s subsequent case on extraterritorial enforcement of the right to be forgotten reflected similar conversations about choice-of-law, jurisdiction, and extraterritoriality from the mid-to-late 1990s. See *CNIL*, *supra* note 34.

the rate of legal change” still resounds throughout law and technology scholarship.³⁷ The pacing problem, where in Reidenberg’s words “today’s regulations may easily pertain to yesterday’s technologies,”³⁸ has led to the development of a robust regulatory toolkit for future-proofing the law.³⁹ That toolkit includes, among other things, deploying technology-neutral versus technology-specific legislation,⁴⁰ debating whether to establish new expert agencies,⁴¹ employing complex forms of hybrid public-private governance;⁴² establishing regulatory sandboxing,⁴³ and using and incorporating extra-legal standards processes.⁴⁴ As Reidenberg wrote back in 1998, many of these regulatory tools evidence, for better or for worse,⁴⁵ “a shift in the focus of government action away from direct regulation and toward indirect influence” on technical development while still attempting to “preserve strong attributes of public oversight.”⁴⁶ We still struggle centrally today with how best to

37. Reidenberg, *supra* note 3, at 566.

38. *Id.* at 586.

39. *See, e.g., Legal Disruption, supra* note 7; Crootof & Ard, *supra* note 7; INNOVATIVE GOVERNANCE MODELS FOR EMERGING TECHNOLOGIES (Gary E. Marchant & Kenneth W. Abbott eds., 2013).

40. Michael Birnhack, *Reverse Engineering Informational Privacy Law*, 15 YALE J.L. & TECH. 24, 27 (2012); Paul Ohm, *The Argument Against Technology-Neutral Surveillance Laws*, 88 TEX. L. REV. 1685 (2010); Brad A. Greenberg, *Rethinking Technology Neutrality*, 100 MINN. L. REV. 1495 (2016).

41. Ryan Calo, *Report: The case for a federal robotics commission*, BROOKINGS (Sept. 15, 2014), <https://www.brookings.edu/research/the-case-for-a-federal-robotics-commission/>.

42. Orly Lobel, *The Renew Deal: The Fall of Regulation and the Rise of Governance in Contemporary Legal Thought*, 89 MINN. L. REV. 342, 346–47 (2004); Margot E. Kaminski, *Binary Governance: Lessons from the GDPR’s Approach to Algorithmic Accountability*, 92 S. CAL. L. REV. 1529, 1530 (2019); W. Nicholson Price II, *Regulating Black-Box Medicine*, 116 MICH. L. REV. 421, 465–71 (2017); Dennis D. Hirsch, *Going Dutch? Collaborative Dutch Privacy Regulation and the Lessons It Holds for U.S. Privacy Law*, 2013 MICH. ST. L. REV. 83, 151–60; Kenneth A. Bamberger & Deirdre K. Mulligan, *Privacy on the Books and on the Ground*, 63 STAN. L. REV. 247, 248 (2011); William McGeveran, *Friending the Privacy Regulators*, 58 ARIZ. L. REV. 959, 980 (2016); Lauren E. Willis, *Performance-Based Consumer Law*, 83 U. CHI. L. REV. 1309, 1330–35 (2015). For critiques, see Cohen, *supra* note 15, at 1915–17; Ari Ezra Waldman, *Privacy, Practice, and Performance*, 110 CALIF. L. REV. (forthcoming 2022) (manuscript at 1, 5–6).

43. Hillary J. Allen, *Regulatory Sandboxes*, 87 GEO. WASH. L. REV. 579, 580 (2019).

44. *See, e.g.,* Emily Bremer, *Incorporation by Reference in an Open-Government Age*, 36 HARV. J.L. & PUB. POL’Y 131, 134 (2013); Irene Kamara, *Co-regulation in EU Personal Data Protection: The Case of Technical Standards and the Privacy by Design Standardisation ‘Mandate’*, 8 EUR. J. L. & TECH., no. 1 (2017).

45. *See* Julie E. Cohen, *How (Not) to Write a Privacy Law*, KNIGHT FIRST AMENDMENT INST. COLUMBIA UNIV. (Mar. 23, 2021), <https://knightcolumbia.org/content/how-not-to-write-a-privacy-law>.

46. Reidenberg, *supra* note 3, at 586.

incorporate expert knowledge and enforcement resources from outside the formal legal system while tethering such lawmaking to the public good.

III. “LEGAL DISRUPTION”

As Reidenberg so presciently laid out in 1998, the interplay between law and technology is complex. Technology can be a regulatory force in its own right, but it does not just act upon static, passive law (or, for that matter, static, passive people).⁴⁷ Rather, the interaction between law and technology takes many forms.

Law can structure technological development, for example, by requiring that technology have certain features (such as seatbelts) or be designed towards certain policy goals (such as safety). Technology can pose challenges for legal institutions qua institutions: by falling outside existing institutions’ fields of specialization or legal mandates; by falling into unclaimed gaps between institutions or regimes; or by falling into a regulatory thicket of overlapping regulation. And of course, there is the so-called “pacing problem,” which usually manifests either as a question of how to get new technical expertise into the legal system or, more generally, how to design law for change over time by delegating some decision-making to more temporally proximate or more expert actors (a variation on the conversation about rules and standards⁴⁸).⁴⁹ As Reidenberg identified in 1998, regulatory design is a perennially central issue for law and technology.⁵⁰

Sometimes, though, the social adoption of a new technology doesn’t raise questions of regulatory design, expertise, or pacing. Sometimes, the adoption and use of technology can make salient existing features of the law.⁵¹ For

47. See, e.g., Jones, *supra* note 7; *Authorship, Disrupted*, *supra* note 7; JULIE E. COHEN, CONFIGURING THE NETWORKED SELF: LAW, CODE, AND THE PLAY OF EVERYDAY PRACTICE 127 (2012) (“[D]eveloping a decentered model of subjectivity organized around three sets of considerations: the evolution of experienced ‘selfhood’ from the situated subject’s perspective, the collective dimension of subjectivity, and the play that overlapping social and cultural networks afford.”).

48. See, e.g., Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 DUKE L.J. 557, 568–69 (1992).

49. See *Legal Disruption*, *supra* note 7, at 14, 21 (listing ways technology and the law interact).

50. See, e.g., Reidenberg, *supra* note 3, at 586–87; *Legal Disruption*, *supra* note 7, at 36 (calling for a legal toolkit for technological change); Crotoof & Ard, *supra* note 7, at 400.

51. Jack M. Balkin, *Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society*, 79 N.Y.U. L. REV. 1, 2 (2004) (“If we assume that a technological development is important to law only if it creates something utterly new, and we can find analogues in the past—as we always can—we are likely to conclude that because the development is not new, it changes nothing important. That is the wrong way to think about

example, the development of participatory online platforms made salient the fact that much of free speech theory and doctrine had been built around a broadcast model of media, with one speaker using legacy media platforms to broadcast to many passive listeners.⁵² This shift also made salient that the law had assumed speech will often be temporary rather than recorded, and limited to certain contexts rather than context-collapsing. Thus, technological change can make salient certain features of doctrine (which doctrinal bucket do internet platforms belong in?⁵³) and of theory (given the cheapness of speech production, is the “marketplace of ideas” theory of the First Amendment outdated?⁵⁴).

As I have argued elsewhere, however, “technology is not just a stable lens through which we see stable aspects of the law. [Technology] takes on a particular meaning within the law depending on what one thinks the law is or should be.”⁵⁵ That is to say, the law dynamically *constructs* technology into its own systems of meaning—just as it constructs many other things.⁵⁶ Take the classic H.L.A. Hart hypothetical “No Vehicles in the Park”: what falls into the doctrinal bucket of “vehicle”? Do cars? Do nonmotorized boats? Do strollers? Do unmanned drones?⁵⁷

technological change and public policy Instead of focusing on novelty, we should focus on salience. What elements of the social world does a new technology make particularly salient that went relatively unnoticed before? What features of human activity or of the human condition does a technological change foreground, emphasize, or problematize? And what are the consequences for human freedom of making this aspect more important, more pervasive, or more central than it was before?”).

52. *Id.* at 6.

53. Oren Bracha & Frank Pasquale, *Federal Search Commission? Access, Fairness, and Accountability in the Law of Search*, 93 CORNELL L. REV. 1149, 1149 (2008); Eugene Volokh & Donald M. Falk, *Google: First Amendment Protection for Search Engine Search Results*, 8 J.L. ECON. & POL’Y 883, 884 (2012); Tim Wu, *Machine Speech*, 161 U. PA. L. REV. 1495, 1496, 1521–22 (2013); James Grimmelman, *Speech Engines*, 98 MINN. L. REV. 868, 870 (2014).

54. Toni M. Massaro & Helen Norton, *Free Speech and Democracy: A Primer for 21st Century Reformers*, 54 U.C. DAVIS L. REV. 1631, 1634–35 (2021); Julie E. Cohen, *Tailoring Election Regulation: The Platform is the Frame*, 4 GEO. TECH. L. REV. 641, 642 (2020); FRANK PASQUALE, *THE AUTOMATED PUBLIC SPHERE* 1–4 (2017). The rise of the sociotechnical phenomenon of online propaganda has led to similar questions about the assumptions behind First Amendment theory and doctrine. See Tim Wu, *Is the First Amendment Obsolete?*, 117 MICH. L. REV. 547, 548 (2018).

55. *Authorship, Disrupted*, *supra*, note 7, at 592.

56. *Id.*; see also Jones, *supra* note 7, at 253 (“[C]yberlaw research should consider the way in which technologies, practices and social arrangements are constructed within certain legal contexts: the legal construction of technology.”).

57. H.L.A. Hart, *Positivism and the Separation of Law and Morals*, 71 HARV. L. REV. 593, 607 (1958); see also Pierre Schlag, *No Vehicles in the Park: Interpretation as Retrieval*, 23 SEATTLE U. L. REV. 381, 387 (1999); Frederick Schauer, *A Critical Guide to Vehicles in the Park*, 83 N.Y.U. L. REV. 1109 (2008).

Sometimes questions of legal construction are just, as Hart tried to argue about his example,⁵⁸ examples of the law operating as usual: applying to tech-enabled social practices without noticeable hiccups;⁵⁹ applying legal definitions to new facts;⁶⁰ establishing “institutional facts”;⁶¹ or situating a fact pattern into existing doctrinal buckets. And sometimes—as is the case with other things the law constructs—things aren’t business as usual, and we have to “bump up a level” to ask why we have a law in the first place.⁶² What might need to change at the level of statute or doctrine in order to accomplish theoretical goals? What might need to change at the level of theory in order to accomplish normative goals? As my colleague Pierre Schlag has written, “[w]e are not just talking about parks and vehicles here; we are talking about parks and vehicles in a legal rule in a legal system in a particular culture.”⁶³

All of these questions—of institutional design, regulatory toolkits, salience, and construction—are ultimately about power. Who has power? How do those with power use it? How does power accumulate, disperse, get checked? And given that we are lawyers, after all, how does the structure of the law exacerbate

58. See Schauer, *supra* note 57, at 1119 (“Hart’s claim, at least in 1958, was that the statutory language, as language, would generate some number of clear or core applications . . .”).

59. PAMELA SAMUELSON, FIVE CHALLENGES FOR REGULATING THE GLOBAL INFORMATION SOCIETY 4 (2000) (“[T]he general view in the U.S. is that antitrust and competition law continues to be viable in the digital age, and can successfully be adapted to deal with software and Internet companies.”).

60. This is what Hart claimed he was doing: merely applying the law—or what Schlag has called “preserv[ing] the hard core of settled meaning from the effects of reconsideration in light of social policy.” Schlag, *supra* note 57, at 387 (observing that Hart’s purportedly authoritative interpretation was a “legal move” like any other).

61. Thanks to the wonderful late Ian Kerr for pointing me to this concept. My favorite example of an institutional fact is a trespass- nuisance case in which a Michigan court defined “dust” as “intangible” for purposes of the law. *Adams v. Cleveland-Cliffs Iron Co.*, 602 N.W.2d 215, 223 (Mich. Ct. App. 1999) (“We further hold that dust must generally be considered intangible and thus not actionable in trespass. We realize, of course, that dust particles are tangible objects in a strict sense that they can be touched and are comprised of physical elements. However, we agree with those authorities that have recognized, for practical purposes, that dust, along with other forms of airborne particulate, does not normally present itself as a significant physical intrusion.”).

62. This resembles Fuller’s response to Hart—you can never just look to the text; you also have to look to the purpose. See Schauer, *supra* note 57, at 1114. I discuss this idea of levels in *Authorship, Disrupted*, *supra* note 7, at 615 (“Examining emergent machine authors and their interface with U.S. law illustrates several ways in which technology can be legally disruptive. Technology can require minor doctrinal tweaks Or it can fall between existing legal categories Or technology can trigger a reassessment of underlying theories behind the law, whether lower level theorization . . . or higher level theorization . . .”).

63. Schlag, *supra* note 57, at 387.

or entrench existing power disparities, or alternatively disperse them and hold power accountable?

IV. “LEGAL DISRUPTION” AND THE IMAGINED REGULATORY SCENE

Technology doesn’t drive the law.⁶⁴ Like other kinds of social changes, however, sociotechnical changes can afford new opportunities for contestation over legal rules and principles.⁶⁵ Whether in courts, regulatory bodies, or legislatures, sociotechnical change often (though not always) creates a chance to reevaluate and argue about not just the application of the law but the normative scaffolding undergirding it. Ostensibly, this is why so many of us remain committed to the study of law and technology.

The second half of this Article now turns to a particular genre of the “legal disruption” discussion that is by now widely prevalent in the law and technology literature but goes largely unidentified and unnamed. We can understand this genre as another example of the “legal construction” of technology, but it is different from the usual debates over textual interpretation or regulatory design. It takes place not on the page but in our heads. It is often the unacknowledged precursor to, or backdrop for, more concrete doctrinal or regulatory conversations.

A growing number of scholars, in a growing number of subfields, have noted that technological adoption can change our fundamental assumptions about the architecture of a regulated environment. Drawing on Jack Balkin and Reva Siegel’s work, I call this move disruption of the “imagined regulatory scene.”⁶⁶

This move has different implications and consequences than discussing how to channel expertise, or regulatory design, or the application of doctrinal buckets—although it may be part of or precursor to any of those conversations, as well. As with any legal move, knowing that it *is* an identifiable move makes a difference in how we understand it.⁶⁷

A. IMAGINED REGULATORY SCENES . . .

Balkin and Siegel identified that every legal principle—by which they mean “norms of conduct that express values”⁶⁸—is developed with a particular imagined paradigmatic scenario in mind. That is:

64. Jones, *supra* note 7, at 253.

65. Balkin & Siegel, *supra* note 9, at 928.

66. *Id.*

67. Schlag, *supra* note 57, at 387–88.

68. Balkin & Siegel, *supra* note 9, at 930.

[L]egal principles are intelligible and normatively authoritative only insofar as they presuppose a set of background understandings about the paradigmatic cases, practices, and areas of social life to which they properly apply. A principle always comes with an *imagined regulatory scene* that makes the meaning of the principle coherent to us.⁶⁹

Balkin and Siegel provide several examples. The imagined regulatory scene behind the First Amendment principle that the government should not discriminate against speech on the basis of its content is government censorship of Communist literature or of antiwar protestors.⁷⁰ The imagined regulatory scene behind the anticlassification principle is Jim Crow, particularly the *de jure* racial segregation of school children.⁷¹

Perhaps these imagined scenes arise naturally out of the common law process, which builds principles from the facts of particular cases. Or perhaps they are endemic to legal reasoning writ large. All words have meaning in context, whether the facts are imagined or applied.

When new circumstances arise that depart from the imagined paradigm, whether through changes in social practices or through technological

69. *Id.* at 928 (emphasis added). The imagined regulatory scene is related but not identical to the concept of “sociotechnical imaginaries” used in Science and Technology Studies (STS). An imagined regulatory scene, as Balkin and Siegel conceive of it, is the more constrained landscape in which construction of a particular law, legal principle, or future legislation or regulation takes place. According to Sheila Jasanoff and Sang-Hyun Kim, sociotechnical imaginaries are “collectively imagined forms of social life and social order reflected in the design and fulfillment of nation-specific scientific and/or technological projects.” Sheila Jasanoff & Sang-Hyun Kim, *Containing the Atom: Sociotechnical Imaginaries and Nuclear Regulation in the U.S. and South Korea*, 47 *MINERVA* 119, 120 (2009). First, sociotechnical imaginaries are often future-oriented in a way an imagined regulatory scene need not be. See Lisa Messeri & Janet Vertesi, *The Greatest Missions Never Flown: Anticipatory Discourse and the “Projectory” in Technological Communities*, 56 *TECH. & CULTURE* 54, 55–56 (2015) (discussing “shared future-oriented narratives about technoscientific possibilities”). Judges and regulators regularly construct imagined regulatory scenes using current or past social practices. Second, sociotechnical imaginaries are quintessentially collectively constructed, on a community or societal level, where an imagined regulatory scene can be individualized (say, at the level of an individual author or judge, obviously influenced by societal-level imaginaries but not necessarily coextensive with them and often in conflict). Third, an imagined regulatory scene is oriented towards answering specific questions about law or regulation in a way that a sociotechnical imaginary need not be. That is, some imagined regulatory scenes might also be sociotechnical imaginaries, but not all; and not all sociotechnical imaginaries are imagined regulatory scenes. Many thanks to Meg Jones, Julie Cohen, and Ryan Calo for pointing out the resemblances, and particular thanks to Meg Jones for providing a guide to the literature.

70. Balkin & Siegal, *supra* note 9, at 931.

71. *Id.*

development or both, a legal principle may become “unstuck.”⁷² Balkin and Siegel also refer to this as “disturb[ing] the ecology of a principle’s application.”⁷³ This creates an opportunity for contestation between actors and counter-actors, not only at courts, but at any lawmaking venue.⁷⁴ Actors might contest whether to apply a particular principle, how to apply a particular principle, or whether the principle remains normatively valid at all.⁷⁵

Balkin and Siegel focus on what happens when there is a complete *shift* from one imagined regulatory scene to another. That is, their focus is on what happens when we shift, for example, from applying the First Amendment doctrinal prohibition against content-discrimination to direct government censorship of disfavored voices, in the first instance, to copyright legislation instead. This is one flavor of legal disruption: technological development and use (filesharing) leads to legal change (copyright legislation) which leads to an entirely different imagined regulatory setting (evaluating the term length of copyright law) for debating the application of foundational legal principles (the First Amendment’s protections).⁷⁶ Through strategic litigation and advocacy, social practices that were once invisible to the First Amendment have now become salient to it—from campaign contributions to computer code to video recording to consumer disclosures.⁷⁷ Balkin and Siegel’s central claim is that shifts in imagined scenes, sociotechnical or otherwise, often unmoor legal principles and make them contestable again.⁷⁸

B. . . . AND ARCHITECTURAL DISRUPTION

The move now recurring in the law and technology literature is different in degree, and perhaps in kind, from the shifts Balkin and Siegel discuss. Rather than asking, “What happens to this old principle when it’s applied in an entirely new setting?” the move asks, “What happens to this old principle when the

72. *Id.* at 928 (“When that background understanding is disturbed the principle becomes ‘unstuck’ from its hermeneutic moorings; it no longer seems clear how the principle applies or even whether it should apply.”).

73. *Id.* at 937.

74. *Id.* at 946.

75. *Id.* at 943–44.

76. *Id.* at 945–46. In the First Amendment context, this question runs parallel to discussions of constitutional salience. See Frederick Schauer, *The Boundaries of the First Amendment: A Preliminary Exploration of Constitutional Salience*, 117 HARV. L. REV. 1765, 1765 (2004).

77. Robert Post, *Encryption Source Code and the First Amendment*, 15 BERKELEY TECH. L.J. 713, 716 (2000); Amanda Shanor, *The New Lochner*, 2016 WIS. L. REV. 133, 140 (2016).

78. Balkin & Siegel, *supra* note 9, at 937 (“[T]echnological change alone did not alter the meaning of the free speech principle; rather, it provided an incentive and an opportunity for interested parties to offer new, competing interpretations of the jurisdiction of the free speech principle.”).

balance of forces in an imagined scene is thrown off?” That is, it incorporates the insights of Reidenberg and others about what technological developments can do to an imagined scene.

This returns us to the core insight of *Lex Informatica*: that technological design can be its own regulatory force. Both Reidenberg and Lawrence Lessig after him understood that technology can be architecture.⁷⁹ Other forces such as laws, market forces, and social norms do matter, but technology is a regulatory force of its own.

That is, technology can constrain behavior,⁸⁰ enable behavior,⁸¹ and mediate behavior.⁸² A website’s design may prevent a user from behaving in particular ways, such as accessing certain material or viewing particular user profiles. A website’s design may enable a user to resort to self-help by deploying privacy-protective technology such as Do Not Track. Or a website’s design may *change the user* by mediating her capabilities and choices.⁸³

The basic point is this: sociotechnical change often alters the imagined regulatory scene’s *architecture*. That is, it changes the imagined paradigmatic scenario not by departing from it entirely, but by constraining, enabling, or mediating behavior, both by actors we want the law to constrain and actors we want the law to protect.

These changes often go beyond the purely architectural. By altering the architecture, sociotechnical change affects social norms, social practices, and social sanctions that the law often presupposes as part of its imagined regulatory scene. If we think of policymaking as being about striking a balance in service of underlying principles, these alterations to the imagined regulatory scene can throw the existing balance out of whack, even as the law on the books remains the same.

To make this all more concrete, let us turn to some illustrations and an example.⁸⁴

When Judge Easterbrook spoke about the relationship between technology and the law, he largely thought of the law as acting upon technology:

79. *The Law of the Horse*, *supra* note 16. LESSIG, *CODE*, *supra* note 16.

80. *See* Reidenberg, *supra* note 3, at 554–55; LESSIG, *CODE*, *supra* note 16.

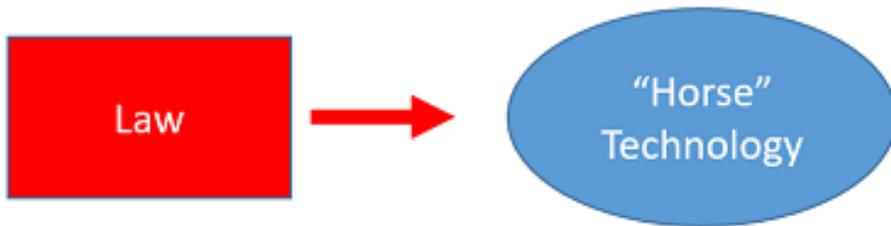
81. *See* Reidenberg, *supra* note 3.

82. *See* Cohen, *supra* note 15, at 1905–06.

83. *See* Calo, *supra* note 15 at 995.

84. I’ve used a variation on this example in talks and in Margot E. Kaminski, *Regulating Real World Surveillance*, 90 WASH. L. REV. 1113, 1136 (2015).

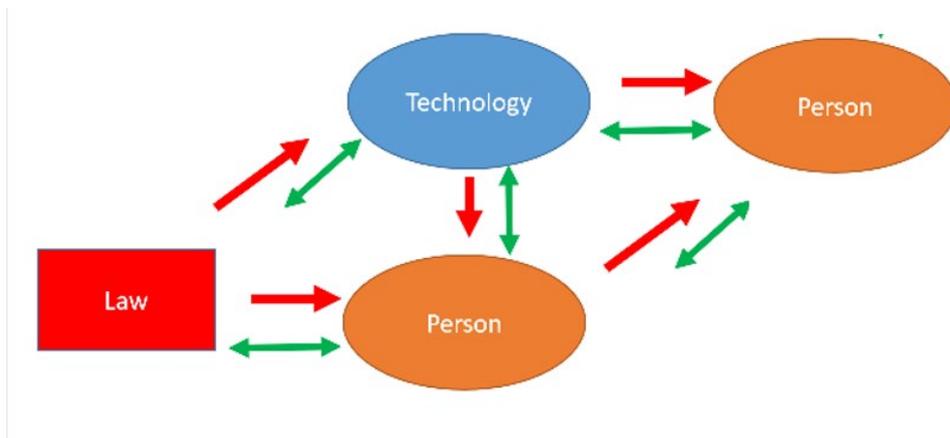
Figure 1: Easterbrook’s Vision



That is, technology is just an object like a horse is an object. The law applies to it, acts upon it, constrains it, or protects it. That is all.

When Reidenberg spoke about technology, however, his conception of the imagined regulatory scene was far more complex. Technology can act upon people just as the law acts upon people. And, importantly, people—and the law—can also act upon, and through, technology:

Figure 2: Reidenberg’s Vision



The above static graphic is woefully inaccurate in that there is no such thing as a core, static “person” or “technology” or “law.” These are not just one-way forces but dynamic, dialectical relationships.⁸⁵ Each is constantly evolving, shaped by and in conversation with the other. And this occurs within

85. COHEN, *supra* note 47, at 131 (“Experienced selfhood is more accurately described as evolving subjectivity, formed and re-formed out of productive tensions between intake and outflow, performance and reflection, contact and separation.”).

culture, where norms shape these interactions and co-evolve with both legal and technological change.⁸⁶

Now for a more concrete illustration: let's take the imagined regulatory scene behind U.S. privacy law's founding document—Warren and Brandeis's *The Right to Privacy*.⁸⁷ Writing in 1890, Warren and Brandeis were concerned about the rise of “instantaneous photographs” and “newspaper enterprise”—a particular technology and a particular social practice.⁸⁸ In 1890, there were no privacy torts; that is, there was no applicable law to either constrain behavior or enable self-help to protect rights. But prior to 1890 (more or less) there were no easy-to-use cameras and no “yellow journalism” or gossip rags to buy, circulate, and profit from scandalous pictures.

Thus, the imagined regulatory scene for Warren and Brandeis initially did not need law to constrain people from recording and circulating private information in the form of personal photographs. For one, prior to the rise of yellow journalism, such information didn't have a market. Without the motivation or means to circulate private information to the general public, much circulation of information could be controlled through social sanction—by, say, socially exiling or shaming the gossip. And—here is technology as architecture—it would be costly, in terms of not just money but time and skill, to sit down and draw a particular person or event from memory, compared to taking a picture.⁸⁹

The *lack* of technology and *lack* of yellow journalism were, in other words, features of some imagined regulatory scene, pre-privacy tort. The balance of forces within that imagined setting achieved a particular policy objective, or served a particular legal principle, without a need for law. Once both the technology and accompanying social practices changed, however, Warren and Brandeis argued that new law was necessary to preserve the policy balance and achieve the same goals.

Now let's look at a second, more contemporary example of almost the same debate. Laws now (somewhat) constrain what people can do with cameras.⁹⁰ People have generally adapted to a world in which ordinary

86. See Jasmine McNealy, *An Ecological Approach to Data Governance* 20–27 (Feb. 20, 2021) (unpublished manuscript) (on file with author) (describing a micro, meso, and macro layered approach to understanding big data, with culture permeating throughout).

87. Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193 (1890).

88. *Id.* at 195.

89. You see much of the same conversation about lowering cost and increased accuracy-reliability in court cases about audio recording. See Kaminski, *supra* note 84, at 1152 n.158.

90. See, e.g., Cal. Civ. Code § 1708.8(b) (West 2011) (regulating recording where a “physical impression could not have been achieved without a trespass unless the visual or

photography is omnipresent. That is, the norms around the pervasiveness of photography have certainly changed, but we have also codified acceptable and unacceptable behavior, including through tools outside of the law, such as social norms and technological architecture.

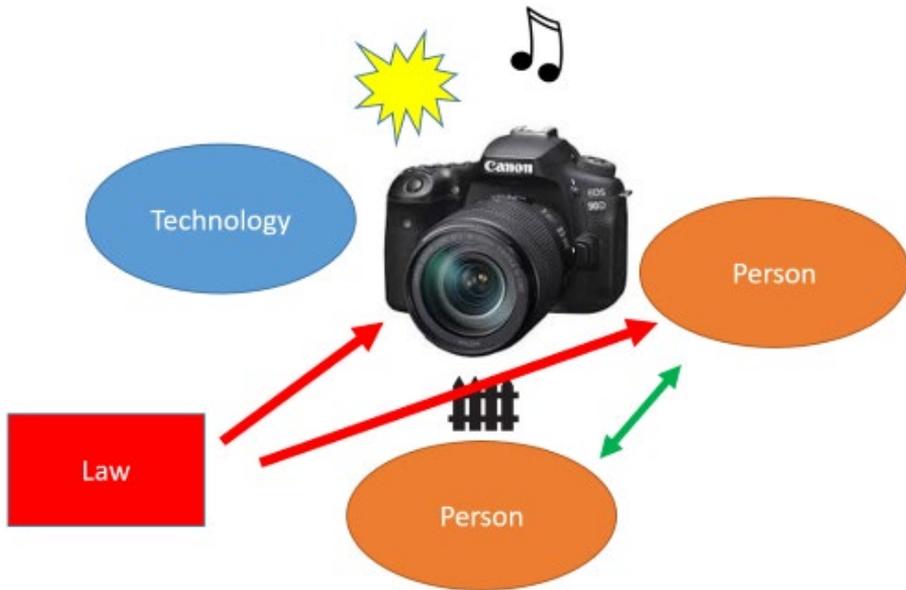
Many cameras are visible and audible, providing notice to picture subjects through shutter sounds or flashes.⁹¹ Often (though definitely not always) people are able to prevent unwanted photographs from being taken. They can socially sanction photographers without resorting to the law. People can and do use the architecture of their lived environments, such as high fences or walls, to keep photographers out. Or they rely on custom and experience to assume that there are no cameras in certain environments, even in ostensibly public spaces.⁹² Figure 3 illustrates how this blend of law, technological design, and social sanction might work to constrain and enable behavior around photography. Law can regulate the technology; law can regulate the person who uses the technology; the person whose image is captured can socially sanction the photographer; and the person whose image is captured can choose to erect a privacy fence or hide behind physical structures:

auditory enhancing device was used”); *Shulman v. Grp. W Prods., Inc.*, 18 Cal. 4th, 200 (1998); *United States v. White*, 401 U.S. 745, 787 (1971) (Harlan, J., dissenting).

91. M. R. Calo, *Against Notice Skepticism in Privacy (and Elsewhere)*, 87 NOTRE DAME L. REV. 1027, 1027 n.1 (2013) (“[A] bill was proposed in the United States that would have required cell phone cameras to make an audible shutter sound. *See* Camera Predator Alert Act of 2009, H.R. 414, 111th Cong. (2009).”).

92. *Rooftop ‘Newlyweds’ Captured by Accidental Drone Shot*, BBC (Oct. 3, 2016), <https://www.bbc.com/news/world-asia-china-37538169>.

Figure 3: Cameras and the Imagined Regulatory Scene



But now cameras can fly. What do drones or unmanned aerial vehicles do to this imagined regulatory scene? I've argued that drones shift the imagined scene in some pretty profound ways.⁹³ Drones may still be noisy and visible, for now but, like hidden cameras, they physically distance the photographer from the photographed. That is, the photographer or videographer often can't be shamed into stopping. Drones, too, change the expected vantage point. Like helicopters or airplanes, they make existing architectural defenses, such as fences, immaterial.⁹⁴ Finally, drone photography or videography, by virtue of its cheapness and ease of adoption (compared to learning to fly or chartering a helicopter), means there is a potential for ubiquitous or pervasive surveillance, which is different in kind even from most existing aerial photography by being different in degree.

Drones, then, arguably alter the law's imagined ecology in a number of ways. The imagined regulatory scene has not shifted to an entirely new environment but features of the imagined environment have changed so as to upset some equilibrium in service of some underlying principle.

93. See Kaminski, *supra* note 84, at 1162.

94. See Surden, *supra* note 18, at 1606 n.3 ("A fence is an example of a structural regulator. Rather than relying upon trespass law to keep unwanted visitors from one's land, landowners often rely on the physical regulation that a tall fence imposes.").

To summarize: technology can alter the imagined regulatory scene. It does so not just by serving as a distinct regulatory force but by upsetting some “balance” of forces within the imagined scene that serves a legal principle. The upset of balance can threaten a legal principle or cause us to reexamine it. Legal responses to these changes, including no response, can shift who has power or can entrench existing power disparities.

C. IMPLICATIONS

As discussed below, identifying how sociotechnical change alters the imagined regulatory environment’s architecture, and thus the balance of legal and normative forces, is an analytical move that now recurs in the legal literature. I am certainly not the first, and will not be the last, to make it. Identifying this as a common move in law and technology analysis lets us better examine its implications.

First, although this move often focuses on constraints on bad actors, it can also undergird conversations about lost or gained affordances and about technological mediation. That is, changes to the imagined regulatory scene affect both constraints on bad actors and the rights and capacities of those we want the legal system to protect.⁹⁵ They also can profoundly affect individual actors by mediating or channeling their behavior.

Take, for example, the ease with which a file can be distributed online. Through the lens of copyright or privacy policy, the change to the imagined scene through the widespread use of this technology lowers constraints on copyright or privacy violators.⁹⁶ Through the lens of the First Amendment, however, this same change is less about constraints and more about capacities: speakers can use the same infrastructure to amplify private voices and more readily become part of public discourse and shared culture.

Technology does not just constrain people. It also provides them tools to exercise their capabilities. Thus, the observation that sociotechnical change has affected the architecture of the imagined regulatory scene is often used to argue for using the law to influence or regulate technological design—for example,

95. While much of Lessig’s work focuses on the latter—on how the various forces of architecture, norms, market, and the law *stop* people from doing something bad—Reidenberg’s work also addresses the former: how architecture not only constrains but *affords*. See Reidenberg, *supra* note 3; see, e.g., Julie E. Cohen, *Affording Fundamental Rights: A Provocation Inspired by Mireille Hildebrandt*, 4 CRITICAL ANALYSIS OF LAW 78 (2017) (reviewing Hildebrandt’s SMART TECHNOLOGIES AND THE END(S) OF LAW (2015)).

96. Surden, *supra* note 18, at 1618 (“Importantly, many emerging technologies possess exactly this characteristic—the tendency to lower transactional and operational costs. This in turn permits conduct which was previously costly or impossible.”).

by reinstating “friction” in online content sharing⁹⁷ or by making privacy-protective features the default or readily available for users to choose.⁹⁸

Sometimes then, a shift in the imagined regulatory scene leads to arguments for replacing pretechnological constraints with new law.⁹⁹ Sometimes it leads to calls to “slow down”¹⁰⁰ or “add friction” or otherwise reimpose old architectural constraints.¹⁰¹ Sometimes it leads to arguments for replacing lost architectural constraints with new technological features.¹⁰² Other times, as in the free speech context, it leads to questions of whether law should celebrate technology’s affordances or restrict them.¹⁰³

More recent literature recognizes that technological architecture also mediates people.¹⁰⁴ That is to say, technological design channels peoples’ behavior, both in what they do and what they watch, hear, or read. It can, often deliberately, lead people to act, buy, or choose things they otherwise would not have done, bought, or chosen—including elected officials.¹⁰⁵ Understanding the ways in which the imagined regulatory scene of, say, news consumption has shifted when it takes place online can be helpful for framing discussions of content moderation and fake news.

The move does not have to be dystopian. It can be applied in more optimistic ways, imagining how the design and use of technology might make the imagined regulatory scene better for the actors in it.¹⁰⁶ And although the

97. William McGeveran, *The Law of Friction*, 2013 U. CHI. LEG. FOR. 15 (2013).

98. Ian Kerr, *The Devil is in the Defaults*, 4 CRITICAL ANALYSIS L 91 (2017).

99. See Surden, *supra* note 18 at 1619.

100. Tom C.W. Lin, *The New Financial Industry*, 65 ALA. L. REV. 567, 603 (2014) (suggesting “slow[ing] down” as a solution).

101. Paul Ohm & Jonathan Frankle, *Desirable Inefficiency*, 70 FLA. L. REV. 777, 804 (2018) (defining “desirable inefficiency” as “fail[ing] to minimize the consumption of time, energy, or space in satisfying a specification of correctness for a given basic problem in order to address a different, related enhanced problem.”).

102. See Woodrow Hartzog & Frederic Stutzman, *Obscurity by Design*, 88 WASH. L. REV. 385 (2013).

103. See Eugene Volokh, *Cheap Speech and What it Will Do*, 104 YALE L.J. 1805 (1995); Wu, *supra* note 54; Massaro & Norton, *supra* note 54.

104. Cohen, *supra* note 15.

105. See, e.g., Jamie Luguri & Lior Jacob Strahilevitz, *Shining a Light on Dark Patterns*, 13 J. LEG. ANALYSIS 43 (2021); Lauren E. Willis, *Deception By Design*, 34 HARV. J. L. & TECH. 115, 143 (2020).

106. See, e.g., Ryan Calo & Danielle Keats Citron, *The Automated Administrative State: A Crisis of Legitimacy*, 70 EMORY L.J. 797, 838–39 (2021) (“Although we decry the actual deployment of automated software systems by agencies to date, we would not deny our government the technological affordances of the twenty-first century. As a diverse set of scholars have begun to observe, agencies can and sometimes do bring advances in information technology constructively to bear on the incredibly complex task of regulation and governance.”).

move is often used to argue for *more* law or *extending* law, it does not necessarily push in a particular normative or regulatory direction. Rather, as Balkin and Siegel note, a shift in the imagined regulatory scene opens up a site of contestation for arguments to occur. Furthermore, different people may have different imagined regulatory scenes in mind, such that an argument to return to or depart from the status quo could point in very different directions.

Counterarguments to making this move include that it is inherently conservative in nature. The move arguably imagines some pretechnological halcyon age of perfect policy balance. Arguably, because of this inherent conservatism, the move can prevent the coevolution of social norms with the uptick of a new technology’s use.

Countermoves include asking whether existing technology or practices in fact have already changed the imagined regulatory scene. If the law has not adapted to cover other similar technologies, then why change it to cover this particular new technology? This “antidiscrimination” argument,¹⁰⁷ which can also come in the guise of an argument for technological neutrality, can be used to push against expanding the law’s coverage or to push for deregulation.

D. THE MOVE IN THE LITERATURE

There are countless examples of the move in the law and technology literature. That is, many articles, knowingly or not, identify how sociotechnical change to the imagined regulatory scene promulgates structural changes and build policy recommendations accordingly. This Section provides just a few of them. I identify both articles where the move is explicit and where the move is implicit in the backdrop. I provide examples that characterize architectural changes to the imagined scene in a variety of ways: as constraints, as affordances, and as mediation.

1. *Architectural Changes as Constraints*

My colleague Harry Surden provides a head-on examination of architectural changes as affecting constraints in his Essay, *Structural Rights in Privacy*.¹⁰⁸ Drawing on Lessig’s work, Surden observes that certain non-legal regulatory mechanisms “restrict or moderate the level of behavior by increasing (or reducing) *costs* of certain activities.”¹⁰⁹ Surden writes of how structural design can act as “*non-legal* regulatory devices,” constraining actors

107. For an example of similar arguments in the First Amendment context, see Felix Wu, *An Anti-Discrimination Theory of the First Amendment* (May. 1, 2021) (unpublished manuscript) (on file with author).

108. Surden, *supra* note 18.

109. *Id.* at 1610.

from particular behaviors.¹¹⁰ Sometimes society “rel[ies] upon a non-legal constraint mechanism to reliably prohibit unwanted behavior *in the place of and as a substitute for an explicit law*.”¹¹¹

Surden distinguishes between intentional structural constraints, such as fences or cryptography,¹¹² and what he calls “latent structural constraints” or “the current technological or physical state of the world.”¹¹³ He points to several activities that at some point were “so costly in terms of resources and effort as to render them effectively impossible to carry out on a widespread basis.”¹¹⁴ His examples of latent structural constraints include DNA sequencing, searching for personal information in paper court records, and copying copyrighted works (before photocopying machines, PCs, and the internet).¹¹⁵

Surden’s use of the move is normative. He claims that societal reliance on latent (as opposed to intentional) structural constraints can indicate the implicit protection of a “constraint-right.”¹¹⁶ When the emergence of new technologies threatens to remove such constraints, Surden suggests that policymakers should replace nonlegal constraints with legal constraints, recreating structural constraints through law.¹¹⁷ Understanding technological change as a shift in the imagined regulatory scene in this way allows Surden to argue that these are not *new* legal rights but, rather, “the *continuation* of a previously existing right.”¹¹⁸

110. *Id.* at 1606 n.7 (“This focus of this Essay is the Hohfeldian negative right—the duty to refrain from a particular behavior . . .”).

111. *Id.* at 1607.

112. *Id.* at 1612.

113. *Id.* at 1613.

114. *Id.*

115. *Id.* at 1613, 1620.

116. *Id.* at 1607.

117. *Id.* at 1609 (“[P]olicymakers should closely examine the implicit privacy interests . . . to expressly determine whether they merit explicit governance by another regulatory device.”); *see also id.* at 1611–12 (“[N]on-legal constraint mechanisms may give rise to relationships between constraints and behaviors that are, in many respects, functionally equivalent to those relationships which give rise to legal rights. In other words, since certain legal rights—negative individual rights—are defined by reference to behaviors that are constrained, it is analytically useful to conceive of the relationship between non-legal constraints and the behaviors that they constrain as creating analogues to legal interests. To the extent that society relies upon a non-legal constraint, such as structure, to inhibit behavior or reliably protect a ‘right’ *in place of or as a substitute for* a legal constraint that would have had to have been enacted to create an explicit legal right, the constraint-rights framework suggests that policymakers should expressly query whether a corresponding rights-like relationship—a constraint-right—has been established.”).

118. *Id.* at 1619 (“Such a distinction becomes important in the public policy debate over protecting privacy interests where the creation of a new privacy right may prove politically more difficult than the protection of an existing right.”).

However, Surden leaves space for policymakers to choose not to preserve a constraint in service of a particular goal.

Orin Kerr is more explicitly normative in his claim that the policy balance struck in the pretechnological imagined regulatory scene is the right one.¹¹⁹ In *An Equilibrium-Adjustment Theory of the Fourth Amendment*, Kerr proposes that courts begin by conceptualizing a balance of police power struck under “the Fourth Amendment at Year Zero, an imaginary time before the introduction of tools both to commit crimes and to catch wrongdoers.”¹²⁰ This is Kerr’s imagined regulatory scene, policing in the time before the development of information technologies.¹²¹

Kerr explains that changes in technology’s social use can upset the policy balance struck in this imagined scene—specifically, the balance of police power versus individual freedom.¹²² According to Kerr,

New tools threaten the privacy/security balance because they enable both cops and robbers to accomplish tasks they couldn’t before, or else to do old tasks more easily or cheaply than before. For criminals trying to commit crimes, new tools mean new ways to commit offenses more easily and more cheaply, or with less risk of being caught than before. . . . Of course, the police use new tools, too. . . . [T]he new tools can expand government power by letting the government collect more information more easily than before.¹²³

Not every technological change will result in an upset of the policy balance. But when the social use of new technology does upset the balance by expanding police power, according to Kerr, judges must and do restore it.¹²⁴ Kerr calls this approach to Fourth Amendment analysis “equilibrium-adjustment.”¹²⁵

Laura Donohue’s work shows that not everyone shares the same idea of the status quo ante. That is, Kerr’s imagined scene at Year Zero affords the police significant power and discretion, especially when surveillance occurs in

119. Orin S. Kerr, *An Equilibrium-Adjustment Theory of the Fourth Amendment*, 125 HARV. L. REV. 476, 484 (2011).

120. *Id.* at 482.

121. *Id.* at 483 (“Year Zero represents an imaginary time, a sort of beginning of the universe for criminal investigations. It is a fiction, of course.”).

122. *Id.*

123. *Id.* at 486.

124. *Id.* at 487 (“[J]udges adjust Fourth Amendment protection to restore the preexisting level of police power.”).

125. *Id.*

public.¹²⁶ Kerr imagines a status quo ante where the rule is, if you're in public, there is no privacy. Donohue by contrast focuses on previous environmental constraints on police power even in public spaces.¹²⁷ That is, she focuses on the tools the police did not have in the past and, consequently, their inability to cheaply and readily track suspects in public. The gap between Kerr's and Donohue's understandings of desirable Fourth Amendment doctrine demonstrates that just because two different scholars both deploy the move does not mean they will arrive at the same normative or doctrinal endpoints.

2. *Architectural Changes as Affordances*

Architectural changes to the imagined regulatory scene may do more than remove constraints on bad actors or decrease the costs of bad behavior. They also alter the affordances of the imagined environment, including for people that the legal system may want to protect. As danah boyd writes:

The design and architecture of environments enable certain types of interaction to occur. Round tables with chairs make chatting with someone easier than classroom-style seating. Even though students can twist around and talk to the person behind them, a typical classroom is designed to encourage everyone to face the teacher. The particular properties or characteristics of an environment can be understood as *affordances* because they make possible—and, in some cases, are used to encourage—certain types of practices, even if they do not determine what practices will unfold. Understanding the affordances of a particular technology or space is important because it sheds light on what people can leverage or resist in achieving their goals.¹²⁸

boyd's initial imagined regulatory scene is a pretechnological physical public space. She writes that the environments shaped by social media differ from physical public spaces in four key ways: persistence, visibility, spreadability, and searchability.¹²⁹ That is, online content persists or endures longer than offline interactions; online content is visible to a larger audience, that is “public

126. Laura Donohue, *The Fourth Amendment in a Digital World*, 71 N.Y.U. ANN. SURVEY AM. L. 533, 2017, n. 35 (noting that “Orin Kerr, in his postulation of the equilibrium theory of the Fourth Amendment, lists as his first rule of the status quo in rule zero: ‘[T]he police are always free to watch suspects in public’”).

127. *Id.* at 558 (“[A]s the collection and analysis of information requires fewer and fewer resources, constraints that previously played a key role in protecting privacy are dropping away.”)

128. DANAH BOYD, *IT’S COMPLICATED: THE SOCIAL LIVES OF NETWORKED TEENS* 10–11 (2014).

129. *Id.* at 12.

by default, [or] private through effort”;¹³⁰ consumers can more readily share online content; and searching for and finding online content is comparably easy. These features are not just removals of constraints on bad actors, as discussed by Surden, Kerr, and Donohue. They allow social media users, good and bad, to use them as capabilities—they *afford*.

These affordances *can* enable bad actors, to be clear. For example, Mary Anne Franks claims that anonymity, amplification, permanence, and publicity all “exacerbate the impact of harassment” online.¹³¹ In particular, Franks’s notion of “amplification” and boyd’s notion of “spreadability” sound in similar notes to the constraints conversation above: they identify a feature of the online environment that makes it easier to harass someone than in the offline physical space—that is, that removes a structural constraint on bad actors.

But a closer reading of Franks shows something else at work. Franks is concerned not just with changes that make it easier for a bad actor to do something bad. She is also concerned with changes that prevent a rights-holder from protecting her rights through doing something good. That is, Franks believes online anonymity makes it “difficult if not impossible for the targets [of harassment online] to engage in self-help” that would have been possible offline.¹³² Offline, a woman could track and shame her harasser. Online, that sort of self-help is much harder.

Franks points, for example, to the kind of context collapse online spaces enable.¹³³ Where in offline spaces the target of harassment could prevent harassment that occurs on the street from impacting her experience in the workplace, in online spaces she no longer has that kind of control. The changed affordances of the online environment alter her capabilities as much as they alter the capabilities of, or constraints on, her harassers.

The affordances-focused take on the imagined regulatory scene thus emphasizes not just constraints on bad behavior; it also focuses on what good actors rely on and use. For example, Woodrow Hartzog and Frederic Stutzman write about the “major structural differences in online and off-line communication,” drawing on boyd’s and Rob Kling’s works, among others.¹³⁴ Hartzog and Stutzman observe that “[w]e utilize a range of cues and physical structures to figure out how we should present ourselves. For example, our

130. *Id.* at 12.

131. Mary Anne Franks, *Unwilling Avatars: Idealism and Discrimination in Cyberspace*, 20 COLUM. J. GENDER & L. 224, 255–56 (2011).

132. *Id.*

133. *Id.*

134. Woodrow Hartzog & Frederic Stutzman, *The Case for Online Obscurity*, 101 CALIF. L. REV. 1, 10 (2013).

understanding of the private nature of a conversation is moderated by the presence of walls and doors.”¹³⁵ Individuals analogously use features of the online environment to establish “private” environments or actively make it difficult for others to find information.¹³⁶

For example, Hartzog and Stutzman argue that courts are mistaken in viewing online privacy through a private-public dichotomy.¹³⁷ They identify that courts in the predigital era recognized a privacy interest in documents that were not entirely secret but were “practically obscure”—that is, where information “was technically available to the public, but could only be found by spending a burdensome and unrealistic amount of time and effort in obtaining it.”¹³⁸ That is, courts recognized that a person with privacy rights could rely on the affordances of paper documents in protecting her privacy. Hartzog and Stutzman suggest that courts today should similarly recognize privacy interests when individuals use features of the online environment to obscure information online.¹³⁹

This brings us back to Reidenberg’s observation that technology and its design can be used to address problems, not just create them. In follow-on work, Hartzog and Stutzman propose ways of implementing “obscurity by design” by deploying technologies individuals can use to actively foster obscurity in the online environment just as they once had in the offline world.¹⁴⁰ That is, understanding that the structure of the imagined regulatory scene online is different than offline, they propose affording analogous structures in the online environment so individuals can continue identity management the way they once did (and largely continue to do) offline.¹⁴¹

135. *Id.* at 7–8.

136. *Id.* at 16 (“[I]ndividuals exert control over the information they disclose by limiting the audience of the disclosure, by bounding the meaning of the disclosure, and by reflexively adapting the disclosure to the site. In social media, where anonymity often violates social norms or site terms, individuals strategically develop techniques that effectively produce obscurity in disclosure. This is not to say that established techniques of privacy management are invalid in these domains, but rather that new techniques that are contextually appropriate emerge so individuals can maintain their expectation of privacy and obscurity.”).

137. *Id.* at 17, 20.

138. *Id.* at 21.

139. *Id.* at 32.

140. Woodrow Hartzog & Frederic Stutzman, *Obscurity by Design*, 88 WASH. L. REV. 385, 402–07 (2013) (recognizing “smart hyperlinks,” privacy settings, search blockers, de-identifying tools, passwords, and encryption as technologies that could implement “obscurity by design”).

141. Hartzog went on to write a book on the role of technological design in both deceiving internet users about the nature of the online environment and enabling active identity management. *See* HARTZOG, *supra* note 29.

Sociotechnical change can, in summary, do more than remove architectural constraints on bad actors. It can alter the imagined regulatory scene such that the tools a person once had—whether physical or social—are significantly changed or no longer there at all. Legislators, regulators, and judges then make decisions as to whether to restore lost affordances, require notice of changed affordances, or constrain affordances through regulation of design.

3. *Architectural Changes as Mediation or Channeling*

References to architectural changes in the imagined regulatory scene can often sound in law and economics, with technological changes and their social uses characterized as a decrease in transaction costs that formerly thwarted bad actors.¹⁴² But not all structural changes to the imagined regulatory scene are significant because they increase or decrease transaction costs for some imaginary rational actor.¹⁴³ The discussion of affordances evidences this; sometimes the architectural change constitutes a new tool or feature that can alter behavioral patterns by channeling them, or removes an existing tool, or triggers miscalibrated behavior by giving off a misleading signal that an environment is other than what it is.

More recently, the law-and-technology literature has turned to the darker side of technological design, focusing on the way in which technology mediates and channels our behavior in ways that go even further than affordances, to the core of our understandings of the self.¹⁴⁴ Take, for example, Ryan Calo’s

142. See Justin Hurwitz, *The Technological Problem of Social Cost: TPRC Draft 1* (Mar. 31, 2016) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2757358 (“[T]echnological change can affect transaction costs, and therefore the law, in predictable ways.”); JUSTIN “GUS” HURWITZ & GEOFFREY A. MANNE, *CLASSICAL LIBERALISM AND THE PROBLEM OF TECHNOLOGICAL CHANGE* 13 (2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3384300 (“[N]ew technology is often developed and adopted precisely because of its effects on transaction costs. But any change in the incidence or level of transaction costs can significantly alter the optimal initial assignment of rights to maximize the likelihood of voluntary exchange. This means that technology may disrupt the structure of the legal institutions necessary to facilitate efficient, welfare-enhancing outcomes.”).

143. Cohen, *supra* note 15, at 1908 (“The self has no autonomous, precultural core, nor could it, because we are born and remain situated within social and cultural contexts. And privacy is not a fixed condition, nor could it be, because the individual’s relationship to social and cultural contexts is dynamic. These realities do not weaken the case for privacy; they strengthen it. But the nature and importance of privacy can be understood only in relation to a very different vision of the self and of the self-society connection.”).

144. Cohen refers to this as “modulation.” See Cohen, *supra* note 15, at 1912 (“Citizens within modulated democracies—citizens who are subject to pervasively distributed surveillance and modulation by powerful commercial and political interests—increasingly will lack the ability to form and pursue meaningful agendas for human flourishing.”).

work *Digital Market Manipulation*.¹⁴⁵ Calo explains that the change to the imagined regulatory scene is that a person now acts *through* technology. That is, “[t]he consumer of the future is a *mediated* consumer—she approaches the marketplace through technology designed by someone else.”¹⁴⁶ The mediated consumer is not necessarily constrained by technological architecture, nor does she use technological architecture’s affordances; she is channeled through them.

This mediation has consequences. Calo claims that by creating a detailed record of consumer behavior that firms use in their design of consumer interfaces, mediation allows for the mass production of cognitive biases and persuasion of a kind previously unknown.¹⁴⁷ That is, “[a] firm with the resources and inclination will be in a position to surface and exploit how consumers tend to deviate from rational decisionmaking on a previously unimaginable scale. Thus, firms will increasingly be in the position to *create* suckers, rather than waiting for one to be born.”¹⁴⁸

Others have more recently focused on the creation of “dark patterns” and other forms of deliberate online manipulation in the marketplace and elsewhere.¹⁴⁹ Julie Cohen discusses mediation in an even more profoundly disrupting sense. For Cohen, we are not just consumers in a marketplace where the mediators actively exploit our inefficiencies or biases. We are by nature socially constructed, and today’s information technologies are designed to affect how we socially construct ourselves.¹⁵⁰

In *What Privacy is For*, Cohen writes:

Like the other artifacts that we use in our daily lives, networked information technologies mediate our relationship to the world around us. Processes of mediation are partly behavioral. The particular design features of our artifacts make some activities seem easier and more natural and others more difficult, and these implicit behavioral templates, or affordances, encourage us to behave in certain ways rather than others. But processes of mediation are also conceptual and heuristic. Our artifacts organize the world for us,

145. Calo, *supra* note 15.

146. *Id.* at 1002.

147. *Id.* at 1006.

148. *Id.* at 1018.

149. See Luguri & Strahilevitz, *supra* note 105; Daniel Susser, Beate Roessler & Helen Nissenbaum, *Online Manipulation: Hidden Influences in a Digital World*, 4 GEO. L. TECH. REV. 1, 26 (2019).

150. See *supra* note 143 and accompanying quote.

subtly shaping the ways that we make sense of it. Over time we come to perceive the world through the lenses that our artifacts create.¹⁵¹

We are not just constrained or released from constraints by new technologies. We become what they channel us into being. The “person” in the diagrams pages ago isn’t some static, separate actor. She is embedded in and designed by technology and its social use as much as she acts upon or is acted upon by it.

V. CONCLUSION

Writing over twenty years ago, Joel Reidenberg identified that technology could change the regulatory environment in ways lawmakers did not yet understand. Technology, too, could itself be deployed to mitigate these changes. Websites and browsers could be designed differently, and law could play a role in that change.

This Article has identified a particular vein of law and technology scholarship in which these observations have played out and developed. Many legal scholars look at technology and the law by imagining a before-and-after: a time before a technology is in use and a time after its use has taken hold. Technological changes to the imagined regulatory scene often affect the architecture of the imagined regulatory environment. Scholars identify these architectural changes and use them to make normative arguments—to keep the law as it is or to change it. Just as with a complete shift in regulatory scene, these changes can shift us up into conversations about, not just the “how” and “what” of law, but also the “why.” I imagine that, for many of us, this is what makes the field of law and technology worthwhile.

151. Cohen, *supra* note 15, at 1912–13.

LEX AI: REVISITING PRIVATE ORDERING BY DESIGN

Niva Elkin-Koren[†] & Karni A. Chagal-Feferkorn^{††}

ABSTRACT

In his seminal paper from 1997, Professor Joel R. Reidenberg articulated a novel governance framework known as “Lex Informatica.” Under the principles of Lex Informatica, norms are no longer shaped by leaders, legislators, or judges but rather by technological capabilities and design choices that grant users the flexibility to shape their own online experience based on their preferences. A quarter century later, a “second generation” of online governance systems has emerged, making use of artificial intelligence: “Lex AI.”

The manner by which Lex AI stirs our behavior or reality (for example by filtering online content it deems inappropriate, advising a judge to refuse defendant’s request for bail, or recommending a certain book and not others) is based on the aggregation of big data and on predictions of the optimal choice for each individual subject to Lex AI. Given its personalized nature, Lex AI may be perceived as a form of private ordering, one that focuses on the individual rather than on the collective, and—at least when its supports voluntary decision-making—grants individuals the ability to execute their own preferences and choices.

Yet, as we explore in this Article, Lex AI bypasses autonomous choice as it is often based on personalization that is conducted for the user and not by the user. As such, it does not neatly fit the definition of private ordering—the process of setting up of social norms by parties involved in the regulated activity.

As a form of public ordering, on the other hand, Lex AI may be viewed as a superior form of collective governance because it bases its decision on the efficient collection and analysis of granular information regarding actual preferences and behavior. As such, it could possibly address one of the major challenges associated with centralized governance: information failure due to limited and outdated data concerning individuals’ actions and appetites.

As this Article shows, however, path dependency, coupled with the reduced opportunity to signal users’ true preferences or to take part in the deliberation of the applicable norms, may render Lex AI a less efficient and less legitimate form of governance than public ordering.

We therefore argue that Lex AI is a *sui generis* type of governance—one which deserves scrutiny by regulators and policymakers. Naturally, the characteristics of Lex AI also offer significant governance advantages. Shaping Lex AI to enhance social welfare, however, may require a fresh way of thinking about these challenges and the public interventions that might address them.

DOI: <https://doi.org/10.15779/Z38F47GV4B>

© 2021 Niva Elkin-Koren & Karni A. Chagal-Feferkorn.

[†] Professor, Tel-Aviv University Faculty of Law; and a Faculty Associate, Berkman Klein Center at Harvard University.

^{††} Scotiabank Postdoctoral Fellow at the AI + Society Initiative, University of Ottawa and the University of Ottawa Centre for Law, Technology and Society.

TABLE OF CONTENTS

I.	INTRODUCTION	916
II.	LEX INFORMATICA AS PRIVATE ORDERING	923
A.	THE LEGAL FRAMEWORKS OF GOVERNANCE.....	924
B.	LEX INFORMATICA AND THE ADVANTAGES OF PRIVATE ORDERING	927
III.	LEX AI AS A FORM OF GOVERNANCE	930
A.	HOW AI CAN BECOME A FORM OF GOVERNANCE	930
B.	LEX AI AS PRIVATE ORDERING	934
1.	<i>Personalization as an Enabler of Private Ordering</i>	934
2.	<i>Does AI Enable Informed Decision-Making?</i>	938
3.	<i>Does Lex AI Even Let Users Choose?</i>	942
4.	<i>Do Lex AI's Choices Reflect Users' Preferences?</i>	945
C.	LEX AI AS A CENTRALIZED GOVERNANCE MODEL.....	948
1.	<i>Beyond Individual Choice</i>	948
2.	<i>How Does Lex AI Decide Users' Best Interests?</i>	950
a)	Path-dependency	952
b)	Optimization function.....	953
c)	Exploration—Exploitation Tradeoffs	954
3.	<i>The Decline of the Deliberative Space</i>	955
IV.	LEX AI AS A <i>SUI GENERIS</i> ORDERING SYSTEM: LEGAL IMPLICATIONS.....	958
V.	CONCLUSION	961

I. INTRODUCTION

In his seminal paper from 1997, Professor Joel R. Reidenberg articulated a novel approach to governance, one in which information technology affordances and constraints steer human behavior.¹ This path breaking approach to governance has given rise to a proliferation of legal scholarship

1. Joel R. Reidenberg, *Lex Informatica: The Formulation of Information Policy Rules through Technology*, 76 TEX. L. REV. 553 (1998). In a similar vein, Lawrence Lessig has coined the term “code is law” to describe how algorithms govern human behavior alongside with the traditional forms of governance of law, social norms, and markets. LAWRENCE LESSIG, CODE: VERSION 2.0 (2006).

seeking to gain a better understanding of the role of technology in governing human behavior.²

Governance is a fuzzy concept, broadly referring to ordering processes that aim to steer and coordinate the behavior of social actors.³ Different frameworks of governance entail different characteristics and hence introduce various advantages as well as potential concerns. Governance by law, for example, directs people's behavior by a set of explicit binding norms that define rights and duties—norms whose violation would result in legal sanctions.⁴ Governance by law may take the form of top-down public ordering, where concepts of right and wrong are collectively determined by the government and are enforced by law (e.g., criminal law). Governance can also take the form of private ordering. Under private ordering, which is perceived by some as a more legitimate and more efficient form of governance than public ordering,⁵ norms are crafted and voluntarily undertaken by the social agents to which they apply (e.g., contract law, which grants parties freedom to shape their desired agreement).⁶ Markets, to take another example of governance frameworks, coordinate and shape the behavior of various economic actors through market mechanisms of supply and demand mediated by price.⁷

Coining the term “Lex Informatica,”⁸ Reidenberg described yet another type of governance: governance through information technology. Under the

2. See, e.g., Helen Nissenbaum, *From Preemption to Circumvention: If Technology Regulates, Why Do We Need Regulation (and Vice Versa)*, 26 BERKELEY TECH. L.J. (2011); JULIE E. COHEN, *CONFIGURING THE NETWORKED SELF: LAW, CODE, AND THE PLAY OF EVERYDAY PRACTICE* (2012); Deirdre K. Mulligan & Kenneth A. Bamberger, *Saving Governance-by-Design*, 106 CALIF. L. REV. 697 (2018).

3. Jeanette Hofmann, Christian Katzenbach & Kirsten Gollatz, *Between Coordination and Regulation: Finding the Governance in Internet Governance*, 19 NEW MEDIA & SOC'Y 1406 (2016).

4. See Lauren Edelman & Marc Galanter, *Law: The Socio-Legal Perspective*, in 13 INT'L ENCYC. SOC. & BEHAV. SCIS. 604–13 (James D. Wright ed., 2015).

5. See *infra* notes 9, 34–48 and accompanying text.

6. Victor P. Goldberg, *The Enforcement of Contracts and Private Ordering*, in HANDBOOK OF NEW INSTITUTIONAL ECONOMICS 491 (2008).

7. Michael J. Sandel, *WHAT MONEY CAN'T BUY: THE MORAL LIMITS OF MARKETS* 89–93 (2012); Ariel Porat, *Changing People's Preferences by the State and the Law*, 22 THEORETICAL INQUIRIES L. 215, 216–17 (2021).

8. Reidenberg, *supra* note 1, at 555. Information technology may shape behavior by creating affordances. For instance, digital platforms' design choices are shaping the way individual users engage with online content. Facebook's “like” and “share” features, for example, enable users to rank content posted by others, offering peer-based credibility to content, and, at the same time, triggering more engagement by posters of content. See Kyle Langvardt, *Regulating Habit Forming Technology*, 88 FORDHAM L. REV. 129 (2019) (criticizing the way platforms use such quantification features to hook users). The greatness of Lex

principles of Lex-Informatica, norms are no longer shaped by leaders, legislators, or judges but rather by technological capabilities and design choices that grant users the flexibility to shape their own online experience based on their preferences. Reidenberg, therefore, perceived information technology as an enabler of bottom-up, private ordering, whereby parties voluntarily undertake the norms that govern their behavior and thus express their choices.⁹ For example, technology enabling different users to choose different content filters mitigated the tension between the one-size-fits-all norm dictated by governance by law and the diversity of speech norms upheld by users.¹⁰

The pace of technological development has been exponential,¹¹ and only a quarter of a century since Reidenberg's groundbreaking proposition of "Lex Informatica," a second generation of algorithms has emerged. Referred to as "artificial intelligence" or "machine learning" (hereafter "AI" or "ML"), the new generation of algorithms, which is capable of "crunching" enormous amounts of data, learning how to independently solve tasks, and reaching decisions,¹² has arguably led to a new form of governance. A form of governance we call "Lex AInformatica" (or "Lex AI" for short).¹³

Informatica, however, is that it allows individuals to use technological means in order to express and execute their own preferences. Yet, information technology can also be applied to restrict individuals' choice. Unlike governance by legal norms, ethics, or market-forces, individuals might be denied a priori of certain choices that would not be technologically feasible. Indeed, Lex Informatica might be used by governments, or by commercial entities, in manners that block certain behaviors using technological means and thus limiting choice altogether.

9. Reflecting a social order which relies on individuals' choice, Lex Informatica might be perceived as a more legitimate form of self-governance. Moreover, since individuals presumably possess better knowledge of their own wants and needs, governance by Lex Informatica was presumably more likely to enable choices that would efficiently maximize individuals' own utility functions (unless stated otherwise, the term "efficiency" is used in this paper in the context of utility maximization). As we discuss in what follows, Lex Informatica was indeed thought of by Reidenberg to be a digital enabler for the same legitimacy and efficiency that bottom-up private ordering could provide. For further discussion on public and private ordering, see *infra* notes 34–54 and accompanying text.

10. Although governments could enforce their local speech rules using technology (for example, blocking certain websites for users within their territory), other internet users could use the same technology to control, for themselves, what content to filter and what to allow. See *infra* notes 58–64 and accompanying text.

11. Daniel Martin Katz, *Quantitative Legal Prediction—or—How I Learned to Stop Worrying and Start Preparing for the Data Driven Future of the Legal Services Industry*, 62 EMORY L. J. 909 (2013).

12. See *infra* notes 67–72 and accompanying text.

13. Although the term "Lex" in Latin means "Rule," JAMES MARWOOD, POCKET OXFORD LATIN DICTIONARY: LATIN-ENGLISH (2012), and is often used in the context of legal rule, this Article's use of it is not strictly limited to the legal context. In accordance with

This Article argues that Lex AI is a very different form of governance compared to Lex Informatica. It shows why Lex AI, which can be easily mistaken for a private ordering form of governance, is in fact closer in nature to public ordering. At the same time, we will show that Lex AI lacks some of the key characteristics at the basis of public ordering, rendering Lex AI a *sue generis* form of governance—one which is not necessarily inferior to other types of governance but whose unique characteristics ought to be fleshed out and carefully considered when making policy choices regarding Lex AI.

Before proceeding with this line of arguments, however, it is necessary to clarify how AI decision-making constitutes a form of governance. First, AI is used in the public sphere for governance purposes. Consider, for example, the algorithms that assist judges in the judicial process or those law enforcement agencies use to identify suspects.¹⁴ But even mundane uses of AI that take in the private sphere are, in fact, a reflection of governance. For example, Netflix makes personalized recommendations to its users by comparing data collected on each of its subscriber's viewing history with the profiles of millions of others to predict each individual's watching preference or how likely they are to try new content.¹⁵ Notably, by defining which content would become available to users, Netflix or other recommendation systems do not simply reflect preferences but may also practically govern public discourse, shape the construction of meaning, and influence political positions and normative perceptions of social actors.¹⁶ Similarly, navigation apps such as Waze do not

its definition of “rule; principle; condition,” the Article’s discussion of “Lex AI” addresses governance by AI which includes, but is not limited to, governance by law.

14. Consider, for example, Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), a decision support tool assisting judges to assess the risk of defendants’ recidivism or Palantir, considered one of the most powerful (and controversial) law enforcement tools in the world. As the “all-seeing-stones” it was named after, Palantir collects personal data on anyone who might have even the slightest relation to the police’s work identifying terrorists and criminals. See Anne Washington, *How to Argue with an Algorithm: Lessons from the COMPAS ProPublica Debate*, 17 COLO. TECH. L.J. 132 (2019); see also Caroline Haskins, *Scars, Tattoos, And License Plates: This Is What Palantir And The LAPD Know About You*, BUZZFEED NEWS (Sept. 29, 2020), <https://www.buzzfeednews.com/article/carolinehaskins1/training-documents-palantir-lapd>.

15. See HOW NETFLIX’S RECOMMENDATION SYSTEM WORKS, <https://help.netflix.com/en/node/100639> (last visited Feb. 4, 2022).

16. Researchers have documented the effect of such systems on political views, including driving radicalization. See José van Dijck, *Facebook and the Engineering of Connectivity: A Multi-Layered Approach to Social Media Platforms*, 19 CONVERGENCE: INT’L J. RSCH. INTO NEW MEDIA TECHS. 141 (2013); Zeynep Tufekci, *YouTube, the Great Radicalizer*, N.Y. TIMES (Mar. 10, 2018), <https://www.nytimes.com/2018/03/10/opinion/Sunday/youtube-politics-radical.html?searchResultPosition=2>.

simply guide any single driver to her selected destination; they also coordinate multiple drivers, thus generating a certain type of social order. As further demonstrated below, the use of AI to support, supplement, or supplant human decision-making may create new norms and systematically direct the behavior of individuals.¹⁷ In that sense, the use of AI introduces the novel governance framework: Lex-AI.

The literature on governance by AI often focuses on governance *of* AI. Scholars are exploring social and legal tools to render AI decision-making more compatible with principles of fairness, due process, and accountability.¹⁸ This is especially so given the nature of AI systems, which are difficult to predict or explain.¹⁹ Scholars have also focused on *who* is governing behavior by using AI,²⁰ raising concerns over the exercise of governing power by unelected private-sector entities (i.e., private governance).²¹

Nonetheless, missing from these discussions is an inquiry into *how* norms are generated and enforced through the proliferation of AI. How does Lex AI measure as a type of governance? Ultimately, to better govern AI and to fully

17. See *infra* notes 73–86 and accompanying text.

18. See, e.g., Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54 (2019) (discussing the role of the private sector in addressing issues of algorithmic bias, transparency, and accountability); David Freeman Engstrom & Daniel E. Ho, *Algorithmic Accountability in the Administrative State*, 37 YALE J. ON REGUL. 800 (2020) (proposing an accountability structure for the use of AI systems in the public sector); Danielle Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. U. L. REV. 1, 1 (2014) (discussing “technological due process” as a means for oversight on algorithmic systems).

19. According to Frank Pasquale, we are now part of a “black box society”—one where hidden algorithms have enormous power to build, or destroy, numerous individual and collective aspects of our lives. See generally FRANK PASQUALE, *THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION* (2015). One of the significant challenges in that context is algorithms’ opaque nature—their outcomes may often not yield a meaningful explanation as to how the system works or why it decided as it did. See, e.g., Richard Warner & Robert H. Sloan, *Making Artificial Intelligence Transparent: Fairness and the Problem of Proxy Variables* (Feb. 16, 2021) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3764131; Maayan Perel & Niva Elkin-Koren, *Accountability in Algorithmic Copyright Enforcement*, 19 STAN. TECH. L. REV. 473 (2016).

20. See, e.g., Ryan Calo & Danielle Citron, *The Automated Administrative State: A Crisis of Legitimacy*, 70 EMORY L.J. 797 (2021); see also Alicia Solow-Niederman, *Administering Artificial Intelligence*, 93 S. CAL. L. REV. 633 (2020) (addressing the power that AI has granted to the private market to de facto regulate human behavior without the proper means to ensure that private companies will “do the right thing” and protect important interests of the public; proposing a regulatory solution to curb the private sector’s control and also the need to protect certain values that might be hindered as a result of this transition).

21. See Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598 (2018).

understand its social implications, we must first ascertain what is *lost in translation* as we increasingly turn to AI in deciding legal matters that affect rights and duties. Does Lex AI govern behavior in the same way as Lex Informatica? Does it offer the same advantages? Does AI introduce any unique challenges as a subject of legal inquiry? What are the implications of deploying Lex AI so broadly in our daily lives?²²

Using the public-private ordering dichotomy as a lens, our analysis demonstrates that Lex AI introduces a new type of governance: one that does not fit neatly under the public-private binary classification and does not necessarily reflect the key advantages or strengths associated with either of them. This is based on two sets of observations:

The first set of observations relates to the view of Lex AI as a type of private ordering. Lex AI facilitates personalization which, at first glance, might be perceived as introducing a superior form of private ordering.²³ Yet, although the customization offered by Lex Informatica is initiated *by* the user, reflecting their choice, personalization by Lex AI is conducted *for* the user by a centralized system. Lex AI recommendations are often based on an optimization function that also weighs considerations the individual does *not* choose, nor do Lex AI recommendations necessarily reflect users' best (predicted) interests. Lex AI may apply top-down decisions based on "big picture" considerations (such as managing congested traffic in the case of Waze) and on the assumption that, owing to its access to big data and analytical tools, "the system knows best."²⁴

22. See Richard Susskind & Daniel Susskind, *Technology Will Replace Many Doctors, Lawyers, and Other Professionals*, HARVARD BUS. REV. (Oct. 11, 2016), <https://hbr.org/2016/10/robots-will-replace-doctors-lawyers-and-other-professionals>. For a forecast on the percentage of actions currently performed by human professionals that could be replaced by automation, see *Automation Potential and Wages for US Jobs*, MCKINSEY GLOB. INST. (Oct. 1, 2018), <https://public.tableau.com/profile/mckinsey.analytics#!/vizhome/AutomationandUSJobs/Technicalpotentialforautomation>.

23. See *infra* Section III.B.1. At first sight, Lex AI continues Lex Informatica's path of bottom-up, private ordering through technology. If Lex Informatica's novelty was its enabling of the customization of choices by individuals, Lex AI's contribution to the tailor-made governance scheme lies in enabling personalization. Personalization, which is made possible through personal data analytics, ostensibly enables more precise tailoring of services, products, and even legal norms to individual user profiles. Both frameworks therefore enable a departure from governance that applies a single norm to all and instead facilitate tailor-made governance for individuals rather than collectives.

24. Nizan G. Packin, *Consumer Finance and AI: The Death of Second Opinions?*, 22 N.Y.U. J. LEGIS. & PUB. POL'Y 101 (2019) (showing that consumers prefer the recommendations of algorithms over those of human experts).

Bypassing autonomous user choice, Lex AI may in fact look more like a distinct type of collective action mediated by algorithm rather than private ordering. The second set of observations thus relates to Lex AI's centralized governance, top-down nature. Lex-AI introduces a powerful means of collecting, accumulating, and analyzing massive amounts of data, as well as making predictions on people's preferences and needs. Utilizing fine-grained data on individuals' behavior, Lex AI can potentially develop and apply shared norms or even personally tailored norms more efficiently.²⁵ Arguably, these new capabilities may address one of the major challenges associated with centralized governance: information failure due to limited and outdated information concerning the actions and proclivities of each individual.²⁶

As this Article elaborates, however, the ways in which Lex AI generates norms and shapes behaviors raise several concerns. One concern is path dependency: a certain choice or past behavior might be factored in by the system, repeatedly, even when it no longer reflects a user's current (and perhaps not even original) desire. Another concern is the lack of participation in crafting the norms.²⁷ Lex AI may not only shrink opportunities to signal users' true preferences but may also lessen the opportunity to take a deliberative part in the shaping of applicable norms.²⁸

Lex AI entails various advantages and opportunities as a governance form.²⁹ At the same time, however, it lacks some of the major advantages of private ordering while it also suffers several limitations as a form of collective

25. See Ariel Porat & Lior Jacob Strahilevitz, *Personalizing Default Rules and Disclosure with Big Data*, 112 MICH. L. REV. 1417 (2014); Omri Ben-Shahar & Ariel Porat, *Personalizing Negligence Law*, 91 N.Y.U. L. REV. 627 (2016).

26. See *Information Failure*, INVESTOPEDIA, https://www.economicsonline.co.uk/Market_failures/Information_failure.html (last visited Aug. 3, 2021) (“[I]nformation failure exists when some, or all, of the participants in an economic exchange do not have perfect knowledge.”).

27. In discussing a hypothetical “good despot”—an “all-seeing” ruler—philosopher John Stuart Mill advocates participation of citizens in shaping the norms which apply to them, to keep them engaged and committed to the common good. JOHN STUART MILL, *CONSIDERATIONS ON REPRESENTATIVE GOVERNMENT* 46 (Kitchener: Batoche Books, 2001) (1861) (“There is no difficulty in showing that the ideally best form of government is that in which the sovereignty, or supreme controlling power in the last resort, is vested in the entire aggregate of the community, every citizen not only having a voice in the exercise of that ultimate sovereignty, but being, at least occasionally, called on to take an actual part in the government by the personal discharge of some public function, local or general.”).

28. Under the deliberative approach, the legitimacy of rulemaking depends on the meaningful participation in the shaping of norms by those who are affected by them. See *infra* notes 52–53 and accompanying text.

29. See *infra* notes 143–146, 154–156 and accompanying text.

action. Accordingly, Lex AI should be thought of as a *sui generis* type of governance, deserving of closer scrutiny by regulators and policymakers.³⁰

The Article proceeds as follows: Part II will provide background context on Lex Informatica and its facilitation of private ordering. Part III will present Lex AI and analyze its similarities and differences compared to types of private ordering. Concluding that Lex AI cannot be fully described as reflecting users' choices, the discussion will then examine whether Lex AI may be perceived as a superior method for collective ordering. Lastly, Part IV will discuss the policy implications of Lex AI as a form of governance that does not fit neatly under the public-private ordering dichotomy. Achieving a desirable version of Lex AI, we argue, requires fresh thinking on the public intervention in governance by Lex AI aimed at promoting societal goals.

II. LEX INFORMATICA AS PRIVATE ORDERING

Like many legal scholars of his generation,³¹ Reidenberg was puzzled by the challenges arising from the transnational environment the internet introduced. The World Wide Web's launch in the early 1990s and subsequent opening up of the internet to commercial traffic ultimately enabled billions of users to directly interact and exchange content across national borders, subjecting such interactions to national laws applying conflicting speech norms.³² On a practical level, although certain content could be considered legitimate in the jurisdiction of the speaker, it may be considered illegitimate according to the national laws applied to the message's audience. On a more theoretical level, this new type of conflict of laws has challenged the legitimacy of public ordering. Although laws of sovereign governments simultaneously applied within national borders to those who consented to the norm (by participating in elections³³) and to those the norm affects (as applied by sovereigns within national borders), the internet led to a disconnect between the two. For instance, the global nature of internet connectivity might expose users of one (more conservative) jurisdiction to the speech norms applied in

30. This Article does not attempt to propose regulatory solutions for AI-based governance. Nor does it center on the question of who ought to oversee regulation affecting the public good or the consequences of entrusting decision-making power to influential private platforms. It explores the implications of shifting from governance by the rule of law and the "if-then" algorithms featured in Lex Informatica to the governance of social relations by data-driven learning algorithms.

31. See, e.g., David R. Johnson & David G. Post, *Law and Borders—the Rise of the Law in Cyberspace*, 48 STAN. L. REV. 1367 (1996).

32. See Reidenberg, *supra* note 1, at 557–560.

33. See *infra* notes 50–53 and accompanying text.

another (more liberal) jurisdiction. Reidenberg demonstrated how Lex Informatica could address such a challenge by enabling online private ordering using technological design. To provide background, this Part briefly introduces the distinction between public ordering and private ordering and then moves to discuss Reidenberg's private-ordering analysis of Lex Informatica.

A. THE LEGAL FRAMEWORKS OF GOVERNANCE

Private ordering, like public ordering, refers to how norms govern human behavior.³⁴ In both instances, norms regulate behavior by defining an outcome (sanction or reward) attached to the factual conditions defined by the norm. Both public and private ordering assume a form of social control in which behavior is governed by norms defining right and wrong.³⁵

Public ordering refers to rulemaking processes the state designs (e.g., through legislature and regulators). Its norms reflect the outcome of collective action mechanisms that public institutions formulate and apply top-down.³⁶ Private ordering, by contrast, concerns bottom-up processes, where the parties are the ones who choose the norms that will govern their behavior.³⁷ Norms are not only selected by the parties themselves but are also shaped and formulated by them through decentralized processes. Private ordering includes some nonbinding norms (e.g., business practices or community social norms),

34. Jorge L. Contreras, *From Private Ordering to Public Law: The Legal Frameworks Governing Standards-Essential Patents*, 30 HARV. J.L. & TECH. 211, 213 (2017) ("The term 'private ordering' refers to the use of rules systems that private actors conceive, observe, and often enforce through extra-legal means."). As used in this Article, private ordering refers to bottom-up processes where norms are undertaken voluntarily by the parties to which the norms apply. But there are many definitions of private ordering. Some definitions emphasize the sharing of governmental regulatory authority with private actors. *See, e.g.*, Steven L. Schwarcz, *Private Ordering*, 97 NW. U. L. REV. 319, 319 (2002). Other definitions pertain to extra-legal norms. *See, e.g.*, Tehila Sagy, *What's So Private About Private Ordering?*, 45 L. & SOC'Y REV. 923 (2011) (providing an overview of private ordering by diamond sellers in New York City, in the kibbutz in Israel, and among ranch owners in California); Elizabeth Sepper, *Gays in the Moralized Marketplace*, 7 ALA. C.R. & C.L. L. REV. 129, 133–34 (2015) (discussing the "ideal" that private ordering would solve disputes over the Affordable Care Act's contraceptives mandate).

35. *See* Eric A. Posner, *Law, Economics, and Inefficient Norms*, 82 U. PA. L. REV. 1697, 1699 (1996). A norm can be understood as a rule that distinguishes desirable and undesirable behavior accompanied by a sanction or reward. Norms can be shaped based on tradition, the wills of a command system, or market powers—which, in general, reflect individuals' choices and constitute a manifestation of private ordering. *See* MICHAEL J. TREBILCOCK, *THE LIMITS OF FREEDOM OF CONTRACT* 1–2 (1993).

36. WILLIAM D. FERGUSON, *THE POLITICAL ECONOMY OF COLLECTIVE ACTION, INEQUALITY, AND DEVELOPMENT* 16–18 (2020).

37. ROBERT C. ELLICKSON, *ORDER WITHOUT LAW: HOW NEIGHBORS SETTLE DISPUTES* 5-9 (1991).

in addition to some binding legal norms (e.g., those shaped by contracts and agreements rather than by a general law).³⁸ The common feature of all such norms is that the individuals to whom they apply craft and accept them.³⁹ Autonomous choice is thus the grounding principle of private ordering, reflecting its normative core.⁴⁰ As such, and assuming there are no market failures,⁴¹ the Law and Economics approach generally perceives private ordering as superior to centralized regulatory regimes.⁴²

Individuals are assumed to possess the most extensive information about their own preferences and how to execute them optimally⁴³ and are thus considered the best guardians of their own interests.⁴⁴ In other words, parties who enter a private exchange are assumed to have an informational advantage over regulatory agencies and legislators acting through collective decision-making processes. Central sovereigns, on the other hand, lack information on individuals' preferences and capabilities when making decisions on how people

38. ELINOR OSTROM, UNDERSTANDING INSTITUTIONAL DIVERSITY 19 (2005).

39. Private ordering may also refer to extralegal systems, where norms are being created and enforced outside the legal regime. See Lisa Bernstein, *Opting out of the Legal System: Extralegal Contractual Relations in the Diamond Industry*, 21 J. LEGAL STUD. 115 (1992); Robert Cooter, *Economic Theories of Legal Liability*, 5 J. ECON. PERSP. 11 (1991); Schwarcz, *supra* note 34, at 324.

40. HANOCH DAGAN & MICHAEL HELLER, THE CHOICE THEORY OF CONTRACT 41–43 (2017).

41. It is assumed that all parties enjoy complete information regarding the transaction, they act voluntarily, and there are no market failures such as monopolies, externalities, or information failures. The economic theory discusses five major market failures: monopolies, public goods, lack of information, externalities, and transaction costs. See MILTON FRIEDMAN, CAPITALISM AND FREEDOM 13 (1962) (arguing that “[t]he possibility of co-ordination through voluntary co-operation rests on the elementary—yet frequently denied—proposition that both parties to an economic transaction benefit from it, *provided the transaction is bi-laterally voluntary and informed.*”).

42. See TREBILCOCK, *supra* note 35, at 10. For a detailed discussion on private ordering's economic advantages over centralized rulemaking, see Niva Elkin-Koren, *Copyrights in Cyberspace—Rights Without Laws?*, 73 CHI. KENT. L. REV. 1155, 1166–72 (1998).

43. Indeed, scholars of behavioral law and economics have demonstrated that individuals may suffer from biases, consequently advocating policies that would debias boundedly rational individuals and enable autonomous decisions. See generally Thomas S. Ulen, *Behavioral Law and Economics: Law, Policy, and Science*, 21 SUP. CT. ECON. REV. 5 (2014) (discussing whether deviations from the predictions of rational choice theory can indeed be corrected and, if so, whether it would be more efficiently achieved through education or through “choice architecture”).

44. See TREBILCOCK, *supra* note 35, at 7 (arguing that, under the economic model, the fact that the contracting parties entered the transaction voluntarily guarantees that such a transaction actually reflects the optimal bargain that benefits them both, in which case, each of the contracting parties presumably knows what their preferences are and express these through their choices in the market).

ought to behave (or, from an economic point of view, what ought to be produced or consumed).⁴⁵ They also face coordination problems when attempting to achieve the desired social ordering (albeit, social ordering shaped using insufficient information).⁴⁶ Lacking the relevant information regarding the impact of norms on each of the parties affected by each rule, central sovereigns are therefore less likely to generate *efficient* rules. Relatedly, private ordering can also increase efficiency by reducing the risk of vulnerability to public-choice distortions,⁴⁷ as well as by lowering transaction costs associated with collecting information about public preferences.⁴⁸

Compared to private ordering, centralized rulemaking processes would likely be much less successful in guaranteeing that the potential costs and benefits associated with the norm will be accurately determined.

The autonomous choice manifested in private ordering also offers moral justification for the enforcement of norms. First, while top-down regulation often takes a one-size-fits-all approach, private ordering leaves room for more diversity and exploration and is therefore capable of tailoring arrangements to changing circumstances and personal tastes. Second, liberal principles justify governmental use of coercive power when all the norm affects are allowed to participate in its creation.⁴⁹ Legitimate governance must therefore reflect the “consent of the governed,” namely those who are subject to a particular rule and those who are affected by the conduct that is the subject of such a rule.⁵⁰

45. *See id.* at 1–2.

46. *See id.*

47. Centralized rulemaking institutions lack the relevant information regarding the impact of rules on all parties affected and, therefore, are unlikely to generate efficient rules. Whereas a decentralized decision-making process guarantees that parties would internalize the impact of the norm on their utility, a centralized process cannot guarantee that all potential benefits and losses will be accurately observed and reflected in the rule.

48. The exercise of choice presumably guarantees that parties would internalize the impact of the norm on their utility. Thus, the voluntary exchange by informed parties reduces the chances of mistakenly assessing public preferences and, consequently, setting the rule inaccurately. *See* Avery Katz, *Taking Private Ordering Seriously*, 144 U. PA. L. REV. 1745 (1996).

49. *See generally* JOHN RAWLS, *POLITICAL LIBERALISM* (1993); Ronald Dworkin, *What is Equality? Part 1*, 10 PHIL. & PUB. AFF. 185 (1981); Ronald Dworkin, *What is Equality? Part 2*, 10 PHIL. & PUB. AFF. 283 (1981); Ronald Dworkin, *What is Equality? Part 3*, 73 IOWA L. REV. 1 (1987).

50. *See, e.g.*, THE DECLARATION OF INDEPENDENCE (U.S. 1776) (“Governments are instituted among Men, deriving their just powers from the consent of the governed . . . it is the Right of the People . . . to institute new Government, laying its foundation on such principles and organizing its powers in such form, as to them shall seem most likely to effect their Safety and Happiness.”); THOMAS HOBBES, *LEVIATHAN* (1651) (asserting that the source of political legitimacy is the sovereign’s ability to protect those who have consented to obey); Stephen Holmes, *Precommitment and The Paradox of Democracy*, in *PASSIONS AND CONSTRAINT*:

Under the liberal view of democracy, such consent could be acquired through participation in voting. Legitimacy would thus stem from the aggregated will of the constituents as reflected in election results, as long as voting rights are provided equally.⁵¹ Under the deliberative approach, it would take more than the right to participate in the collective decision-making processes to establish legitimacy.⁵² The deliberative approach demands the opportunity to meaningfully participate in the deliberative processes that generate norms.⁵³ Under private ordering, individuals not only give their consent to the shaping of norms but also actively participate in creating them.

The distinction between public and private was often instrumental in giving deference to private-ordering regimes and keeping governmental intervention to the bare minimum, given the advantages of private ordering in terms of legitimacy and efficiency.⁵⁴ As Lex AI introduces new practices that affect individual will and reflect individual and collective choices (as discussed in Part III), it may challenge some of the fundamental tenets underlying liberal ideologies.

B. LEX INFORMATICA AND THE ADVANTAGES OF PRIVATE ORDERING

Lex Informatica offered a conceptual framework for articulating design as a type of private ordering. Reidenberg analogized Lex Informatica to the “Lex Mercatoria” of the Middle Ages. Lex Mercatoria, or the “Law of the Merchant,” was a body of customs and practices merchants developed in the thirteenth century to overcome the lack of a unified legal regime among traders coming from different jurisdictions.⁵⁵ Voluntarily developing a legal framework of their own, merchants were able to gain legal certainty and fairness in an otherwise lawless world.⁵⁶ In other words, Lex Mercatoria was a

ON THE THEORY OF LIBERAL DEMOCRACY 134–77 (1995) (analyzing the foundational assumption that nations’ legality rest on the consent of the governed and how it reconciles with a constitutional convention).

51. Jürgen Habermas, *Three Normative Models of Democracy*, 1 CONSTELLATIONS: INT’L J. CRITICAL & DEMOCRATIC THEORY 1, 3 (1994).

52. See José Luis Martí, *Pluralism and Consensus in Deliberative Democracy*, 20 CRITICAL REV. INT’L SOC. & POL. PHIL. 556, 558–59 (2017).

53. See Stephen M. Feldman, *The Persistence of Power and the Struggle for Dialogic Standards in Postmodern Constitutional Jurisprudence: Michelman, Habermas, and Civic Republicanism*, 81 GEO. L.J. 2243, 2245 (1993); see also Bernard Manin, *On Legitimacy and Deliberation*, 15 POL. THEORY 338, 352 (Elly Stein & Jayne Mansbridge trans., 1987).

54. See, e.g., Schwarcz, *supra* note 34, at 320–21; Lisa Bernstein, *Private Commercial Law*, in 3 THE NEW PALGRAVE DICTIONARY OF ECONOMICS AND THE LAW 108 (Peter Newman ed., 1998).

55. Reidenberg, *supra* note 1, at 553–554.

56. See generally HOBBS, *supra* note 50.

classic example of private ordering where private entities or individuals—not the sovereign—voluntarily shaped and enforced norms.⁵⁷

Nearly half a millennium later, internet users subject to multiple national rules face the same challenges as the merchants of the Middle Ages. Similar to *Lex Mercatoria*, the ground rules that govern the internet and many of the interactions occurring throughout the Web are not set by a national ruler or its public institutions. Rather, Reidenberg observed a new governance framework, which he dubbed “*Lex Informatica*.” For *Lex Informatica*, norms were no longer shaped by leaders, legislators, or judges but rather by technological capabilities and system-design choices that granted users the flexibility to shape their online experience based on their personal preferences.

One prime example Reidenberg analyzed to demonstrate this point was Platforms for Internet Content Selection (PICS), a technical solution that facilitated selective blocking of access to information. PICS, which is no longer used, was a standard for labeling content on the internet that enabled users to instigate automated blocking of websites based on their preferences. Its various components started with the labeling of content by third parties. Here, the system was neutral with respect to the terms used in rating labels; it merely offered a standardized format for rating materials available on the internet.⁵⁸ Such ratings could relate to aspects such as violence, nudity, or adult language;⁵⁹ and, for each dimension, it could assign any number of values (e.g., nudity 1–10). Ratings could be based on self-labeling by the content-provider or by third parties such as the Internet Content Rating Association⁶⁰ or a software-blocking provider running its own rating service.⁶¹

A second component of PICS was a system that automatically detected the rated content and enabled the access blocking. Such rule-based algorithms could be embedded in browsers⁶² or in stand-alone blocking software. The

57. Reidenberg, *supra* note 1, at 553–554.

58. This labeling system has been discontinued by the Family Online Safety Institute’s board of directors. See ICRA, Fam. Online Safety Inst., <http://www.icra.org/> (last visited Feb. 2, 2022).

59. Jonathan Weinberg, *Rating the Net*, 19 HASTINGS COMM’N & ENT. L.J. 453, 457 (1997).

60. *About ICRA*, Internet Content Rating Ass’n, <https://itp.cdn.icann.org/en/files/registry-agreements/net/about-icra-05jan07-en.pdf> (last visited Jul. 25, 2021).

61. Weinberg, *supra* note 59, at 455. Weinberg warned that “[p]eople whose image of the net is mediated through blocking software will miss out on worthwhile speech through deliberate exclusion, through inaccuracies in labeling inherent to the filtering process, and through the restriction of unrated sites.” *Id.* Weinberg compares the different rating strategies of blocking software Cyber Patrol, Specs for Kids, and CYBERSitter. *Id.* at 458.

62. Reidenberg, *supra* note 1, at 559–560.

filters using the PICS specifications were provided as a readymade blocking software that predefined the type of content to be blocked using a filter (e.g., violence with a rating higher than 3); or, alternatively, it offered users a choice to customize their filters based on their preferences (e.g., nudity 0 and violence 5).⁶³

The rule-based technology at the heart of PICS allowed users to select the setting that suited them best from a menu of predetermined options. Although PICS could be used “top-down” by the sovereign,⁶⁴ it was also a tool that enabled individuals to shape their online environment as they saw fit. Individuals could, for example, choose a specific label (or combination of labels) that indicated the content they wished not to see and use PICS to block it from their screens. As a result, individuals “got what they wanted”: content that was offensive to some could be voluntarily blocked by them while others’ wish to continue watching it was respected. In this way, Lex Informatica facilitated customization, flexibility, and multiplicity of norms, resulting in more efficient and more legitimate outcomes.

Reidenberg thus saw the great potential of technological solutions, such as PICS, in enabling the customization and self-enforcement of norms by extra-legal tools. PICS, he argued, allowed individuals to choose their own filtering rules and, at the same time, enabled automated transborder enforcement of norms by content providers without requiring any law enforcement efforts.⁶⁵

Internet Explorer 3 was one of the early web browsers to offer support for PICS.]

63. See Paul Resnick & James Miller, *PICS: Internet Access Controls Without Censorship*, 39 *Comm’n ACM* 87 (1996), <https://doi.org/10.1145/236156.236175>.

64. Reidenberg also demonstrated how Lex Informatica was an “extra-legal instrument” that could enable governments to overcome the jurisdictional challenges posed by a global network. He explained that the law applied a single standard regarding blocking, one that was often different across jurisdictions. Although in the United States the First Amendment would prevent the government from any interference in free speech, in China the law would allow the government to restrict speech of various types. In the United States, in particular, enabling pluralism is the underlying principle of the liberal view of free expression, and the constitutional shield from governmental intervention is designed to ensure sufficient space for such private expression of speech norms. By enabling technological customization through technological solutions such as PICS, Lex Informatica allows governments to create artificial online zones or online jurisdictions where the geographical jurisdiction’s rules are respected. Reidenberg, *supra* note 1, at 557–60.

65. See *id.* at 560 (“The structure of PICS allows several different content-evaluation standards to be applied to the same information on a web site and different viewers to use different filter criteria This technology provides individual choice of filtering rules, yet it still offers automatic enforcement Third-party rating labels may be distributed through a server that is separate from the labelled documents. Thus, the document authors and web sites where the documents are posted need not cooperate with law enforcement efforts.”).

III. LEX AI AS A FORM OF GOVERNANCE

A. HOW AI CAN BECOME A FORM OF GOVERNANCE

As this Article explores in depth, AI as a form of governance introduces several unique traits that warrant careful consideration. To better understand them and how AI can be used for governance purposes, this Section reviews how AI systems works.

AI systems use algorithms and data to identify patterns and make predictions. Despite the multitude of conceptualizations proposed for the term “AI”—or perhaps because of it—there is no consensus over its definition.⁶⁶ AI is generally used to broadly describe different types of techniques. A general common denominator of AI systems could be their ability to replace human beings in performing actions that typically require human cognitive skills.⁶⁷ When further deconstructed, the ability to replace humans may focus on the systems’ ability to act in a manner that is not preprogrammed and to adapt their actions to the changing environment.⁶⁸ Machine Learning (ML) techniques, among others, enable systems to learn how to perform a certain task⁶⁹ by training on vast volumes of data. Once trained, these systems enter

66. Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, 51 U.C. DAVIS L. REV. 399, 404 (2017) (arguing that “[t]here is no straightforward, consensus definition of artificial intelligence”); Bryan Casey & Mark Lemley, *You Might Be a Robot*, 105 CORNELL L. REV. 287, 294 (2019) (arguing that “[t]here is something exceptional about robots and AI that make them exceptionally difficult to define”).

67. *Id.* at 404 (explaining that “AI is best understood as a set of techniques aimed at approximating some aspect of human or animal cognition using machines”).

68. See, e.g., *Annex to the Resolution: Recommendations as to the Content of the Proposal Requested*, EUR. PARL. DOC. P8_TA(2017)0051, ¶ AB, <http://www.europarl.europa.eu/sides/getDoc.do?type=TA&reference=P8-TA-2017-0051&language=EN&ring=A8-2017> (“A common European definition for smart autonomous robots should be established, where appropriate including definitions of its subcategories, taking into consideration the following characteristics: the capacity to acquire autonomy through sensors and/or by exchanging data with its environment (inter-connectivity) and the analysis of those data; . . .”); ANNONI ALESSANDRO, PETER BENCZUR, PAOLO BERTOLDI, BLAGOJ DELIPETREV, GIUDITTA DE PRATO, CLAUDIO FEIJOO, ENRIQUE FERNANDEZ MACIAS, EMILIA GUTIERREZ, MARIA IGLESIAS PORTELA, HENRIK JUNKLEWITZ, MONTSERRAT LOPEZ COBO, BERTIN MARTENS, SUSANA FIGUEIREDO DO NASCIMENTO, STEFANO NATIVI, ALEXANDRE POLVORA, JOSE IGNACIO SANCHEZ MARTIN, SONGUL TOLAN, ILKKA TUOMI & LUCIA VESNIC ALUJEVIC, *ARTIFICIAL INTELLIGENCE: A EUROPEAN PERSPECTIVE* 8 (Max Craglia ed., 2018).

69. In the past, there were attempts to develop AI systems with “general human intelligence,” referred to as “strong AI.” Recent years’ focus, however, has been on developing AI systems capable of replacing humans in concrete tasks or narrow domains (“Narrow-AI”), or on systems capable of accomplishing several tasks in various domains (“Artificial General Intelligence”). See Stan Franklin, *History, Motivation and Core Themes*, in *THE CAMBRIDGE*

an organic process of continual learning that relies on a recursive feedback loop. AI systems that are deployed in governing human behavior eventually attain the capacity to analyze individual personal data by drawing on their prior learnings to make predictions about individuals' preferences,⁷⁰ the risk these individuals may pose,⁷¹ or the opportunities associated with them.⁷²

Similar to Lex Informatica, AI decision-making could be perceived as a form of governance in terms of how it generates norms, shapes practices, and coordinates the behavior of social actors.⁷³ This broad view of governance is not limited to command-and-control by state agencies but rather covers a whole range of regulatory interventions by various social actors. As a governance tool, governments and the private sector alike can use Lex AI. In the public sphere, AI systems are used in various contexts such as smart cities,⁷⁴ welfare benefits, education, and immigration.⁷⁵ In the criminal justice realm, algorithms are often deployed as decision-support systems that assist the police in identifying potential suspects, sometimes even before a crime has been

HANDBOOK OF ARTIFICIAL INTELLIGENCE 15, 15 (Keith Frankish & William M. Ramsey eds., 2014).

70. For some nontrivial examples, see Daniel Faggella, *AI in Taste and Art – The Current State of Machine Learning for Understanding Preferences*, EMERJ (Jan. 29, 2019), <https://emerj.com/editorial-opinion/ai-taste-art-current-state-machine-learning-understanding-preferences/>; Kyle Wiggers, *AI Predicts Office Workers' Room Temperature Preferences*, VENTUREBEAT (Mar. 22, 2019), <https://venturebeat.com/2019/03/22/ai-predicts-office-workers-room-temperature-preferences/>.

71. Be it the risk to society, the risk of activating insurance coverage, or, for example, the risk of dying. See Carolyn McKay, *Predicting Risk in Criminal Procedure: Actuarial Tools, Algorithms, AI and Judicial Decision-Making*, 32 CURRENT ISSUES IN CRIM. JUST. 22 (2019); Joachim Frick & Iris M. Barsan, *InsurTech - Opportunities and Legal Challenges for the Insurance Industry*, REVUE TRIMESTRIELLE DE DROIT FINANCIER 56 (2020); Mohammad Pourhomayoun & Mahdi Shakibi, *Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making*, SMART HEALTH, Apr. 2021, at 1.

72. Etinder Singh & Jyoti Doval, *Artificial Intelligence and HR: Remarkable Opportunities, Hesitant Partners*, 4 PROCS. NAT'L HR CONF. ON HUM. RES. MGMT. PRACS. & TRENDS 97 (2019), <https://ssrn.com/abstract=3553448>.

73. David Levi-Faur, *Regulation & Regulatory Governance*, in HANDBOOK ON THE POLITICS OF REGULATION 1–20 (2011).

74. The term “smart cities” refer to cities that are monitored by ubiquitous computer systems and, at the same time, are driven by innovation and entrepreneurship. Rob Kitchin, *The Real-Time City? Big Data and Smart Urbanism*, 79 GEOJOURNAL 1, 1 (2014), <https://ssrn.com/abstract=2289141>.

75. See Robert Brauneis & Ellen P. Goodman, *Algorithmic Transparency for the Smart City*, 20 YALE J.L. & TECH. 103 (2018); David Freeman Engstrom, Daniel E. Ho, Catherine M. Sharkey & Mariano-Florentino Cuéllar, *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies* (2020), <https://www-cdn.law.stanford.edu/wp-content/uploads/2020/02/ACUS-AI-Report.pdf>.

committed (predictive policing),⁷⁶ or a judge in determining the potential risks of offenders in sentencing and release decisions.⁷⁷ The determinations of judicial and semi-judicial decisions regarding social welfare benefits,⁷⁸ fraud detection,⁷⁹ counterfeit products,⁸⁰ and copyright infringement, for instance, are becoming increasingly automated using AI systems.⁸¹ Indeed, as acknowledged by many scholars and policymakers, AI holds great potential for enhancing the administrative state's governance efficiency.⁸²

As discussed in Section II.B, Lex AI might also shape norms when used by private entities. Content moderation is one example: ML systems installed in social media platforms' upload filters detect illicit speech, such as hate speech, terrorist propaganda, and copyright infringements.⁸³ In this context, Lex AI defines the scope of permissible speech by creating speech affordances—that is, determining which content remains available and which content is removed. Through its technical definitions of particular features and their respective weights, Lex AI effectively defines whether a certain piece of content, be it image, text, or video, would be classified as illegitimate speech that is subject to removal. Once content is tagged as hate speech or terrorist

76. Palantir, for example, assists the police in identifying potential terrorists and criminals. *See supra* note 14.

77. COMPAS, for example, is a decision-support tool assisting judges to assess the risk of defendants' recidivism. *See supra* note 14.

78. *See, e.g.,* Alexandra Chouldechova, Diana Benavides-Prado, Oleksandr Fialko & Rhema Vaithianathan, *A Case Study of Algorithm-Assisted Decision Making in Child Maltreatment Hotline Screening Decisions*, 81 *Procs. Machine Learning Rsch.* 1, 1 (2018).

79. *See* MEREDITH WHITTAKER, KATE CRAWFORD, ROEL DOBBE, GENEVIEVE FRIED, ELIZABETH KAZIUNAS, VAROON MATHUR, SARAH MYERS WEST, RASHIDA RICHARDSON, JASON SCHULTZ & OSCAR SCHWARTZ, *AI NOW REPORT 2018 10* (2018). For instance, U.S. federal agencies such as the Securities and Exchange Commission and the Internal Revenue Service are using AI to detect fraud activities. *See* ENGSTROM ET AL., *supra* note 75, at 22.

80. *See Project Zero leverages the combined strengths of Amazon and brands to drive counterfeits to zero*, AMAZON, <https://brandservices.amazon.com/projectzero> (last visited Feb. 2, 2022).

81. *See YouTube Operations Guide: Using Content ID*, YOUTUBE HELP, <https://support.google.com/youtube/answer/3244015?hl=en> (last visited Feb. 2, 2022).

82. For instance, a major report commissioned by the Administrative Conference of the United States argues that AI promises to transform how government agencies do their work, “reduce the cost of core governance functions, improve the quality of decisions, and unleash the power of administrative data, thereby making government performance more efficient and effective.” ENGSTROM ET AL., *supra* note 75, at 6.

83. KIRSTEN GOLLATZ, FELIX BEER & CHRISTIAN KATZENBACH, *THE TURN TO ARTIFICIAL INTELLIGENCE IN GOVERNING COMMUNICATION ONLINE*, HUMBOLDT INST. INTERNET & SOC'Y (2018), <https://www.hiig.de/wp-content/uploads/2018/09/Workshop-Report-2018-Turn-to-AI.pdf>.

propaganda it may not only become unavailable but also be considered illegitimate.⁸⁴

In that sense, AI makes decisions that shape users' behaviors and the public discourse in two ways. First, and unlike Lex Informatica, it does not merely enforce a human-made decision on what constitutes legitimate content. It actually reaches this decision (or shapes this norm) on its own. Second, as further discussed in Part III, the recursive nature of the AI decision-making process might lead to scenarios where an AI's decision on a specific content's legitimacy will later expand to other types of content as well. If, for example, the system decided to classify content *A* as illegitimate and content *B* is similar to content *A*, then the decision regarding content *A*, be it justifiable or not, increases the chances of content *B* also being removed (followed by content *C* and *D* and *E*, etc.). Another example reflecting Lex AI's role as a governance system used in the private sector is Waze, a crowd-sourced navigation app owned by Google. Waze recommends the fastest driving route between two points by integrating users' inputs, tracking their location, and aggregating and processing their self-reports on traffic conditions. To do this, Waze has to calculate the behavior of all drivers who are using the app. Consider, for instance, a traffic jam that requires drivers to be redirected to alternative routes. If all drivers using the app were directed to take the shortest route, it would likely also become congested, and the app would no longer serve the best interest of any of its users. Therefore, theoretically, once all drivers are using Waze, it may need to reach a social optimum for all, even if that runs contrary to the optimum for a single driver. This puts an app in an interesting position. While market pressures may require it to maximize its utility for each individual user (as, otherwise, they might switch to a rival, assuming there is market competition), it would perform better if it maximized its overall utility for *all* users—in effect, assuming the perspective of a traffic controller. In this capacity, it would be *governing* the traffic.

Another manifestation—or, more accurately, consequence—of Lex AI as a governance tool is that it may cause externalities. Maximizing utility for drivers using Waze, for instance, may produce externalities for nonusers of the app. If, for example, Waze diverts the traffic from the jammed freeway to some quiet residential streets, it may negatively affect the wellbeing of those residents, including residents who do not use Waze. To reflect the *common good*,

84. Niva Elkin-Koren & Maayan Perel, *Democratic Contestation by Design: Speech Governance by AI and How to Fix It*, Fla. St. U.L. Rev. (forthcoming 2023)..

the app would also have to take their interests into account when managing traffic patterns and volumes.⁸⁵

Finally, Lex AI could work in collaboration with governments. For instance, Waze could be used to assist in enforcing the law (e.g., speeding limits) by blocking some routes for security or safety purposes (e.g., wildfires in Los Angeles⁸⁶). Police might also use this app, for example, in order to determine where it is needed to direct the traffic.

All in all, Lex AI introduces a new type of governance that facilitates an adaptive and dynamic decision-making process for governing behavior in both the public and private spheres. The next Section discusses whether Lex AI's distinct governance features satisfy the common assumptions regarding private ordering.

B. LEX AI AS PRIVATE ORDERING

1. *Personalization as an Enabler of Private Ordering*

A major feature of Lex AI that characterizes how it governs behavior is its unparalleled capability to enable personalization.⁸⁷ Personalization is made possible through personal data analytics, which enable services to be tailored to particular users' profiles. For example, in the context of content filtering, an AI content-moderation system might learn the ages of a household's children based on data harvested from the parents' social media or other types of information records;⁸⁸ the AI will automatically select the level of filtering. Filtering could be based on how parents with similar profiles have filtered content for children of similar ages. From such data, the system can then

85. See Oren Dori, *What Happens When Waze Becomes Israel's Traffic Cop*, HAARETZ (Nov. 2, 2017), <https://www.haaretz.com/israel-news/what-happens-when-waze-becomes-israels-traffic-cop-1.5462367> (arguing that Waze could turn a side street into a freeway and that it prioritizes individual users over the common good).

86. Alasdair Wilkins, *How Do Navigation Apps Handle the Los Angeles Wildfires? We Asked Waze*, INVERSE (Aug 12, 2017), <https://www.inverse.com/article/39232-california-fires-navigation-app-waze>.

87. Today's internet is breathtakingly personalized and equipped with armies of algorithms fed by unimaginably vast data archives curating the digital world to feed us individualized streams of content designed to maximize our responsiveness, engagement, content production and other monetizable behaviors. Kaleb Leetaru, *Could Personalized Content Moderation Be The Future Of Healthy Social Media?*, FORBES (July 28, 2019), <https://www.forbes.com/sites/kalebleetaru/2019/07/28/could-personalized-content-moderation-be-the-future-of-healthy-social-media/?sh=1dba08e64b53>.

88. See Adi Robertson, *A Facebook patent would use your family photos to target ads: But it can already figure out a lot of details without them*, THE VERGE (Nov. 15, 2018), <https://www.theverge.com/2018/11/15/8096724/facebook-photo-family-demographics-data-mining-patent-application>.

recommend, or de facto execute, the filtering instructions it deems most appropriate for the family.

Among other areas, AI-driven personalized decision-making may pertain to goods and services. For instance, an AI system could personalize a music app's music suggestions based on past listening habits; or, for a dating app, it could personalize dating matches to fulfill each user's specific characteristics and preferences; and it could tailor a language course to a student's particular learning capabilities.⁸⁹ Netflix, for instance, is able to tailor its offerings to particular users based on their previous views rather than broadcasting "standard" content to a wide audience. Google search is another classic example of AI-tailored search results based on user search history.⁹⁰ The collection of data and data analytics also allow suppliers to personalize the price of products and services advertised to consumers.⁹¹ Personalized AI decision-making may also pertain to legal rights. For instance, data collected by firms on the risk level of individuals could be used to complete an incomplete contract according to the individual's assessed level of risk.⁹²

The personalization capability that enables the tailoring of private-sector services to individuals may have a similar effect in the public sphere. Lex AI introduces unprecedented opportunities for personalizing the enforcement and even formulation of norms that govern behavior. In recent years, a growing number of scholars have pointed out the advantages and challenges personalized laws introduce.⁹³ Such systems deployed in legal settings could

89. See, e.g., *Listening is everything*, <https://www.spotify.com/us/> (last visited Feb. 2, 2022) (music); *Find the people you've crossed paths with*, <https://www.happn.com/en/> (last visited Feb. 2, 2022) (dating); DUOLINGO, <https://www.duolingo.com/> (last visited Feb. 2, 2022) (language learning).

90. TARLETON GILLESPIE, *CUSTODIANS OF THE INTERNET* 195 (2018).

91. See Pascale Chapdelaine, *Algorithmic Personalized Pricing*, 17 N.Y.U. J.L. & BUS.1 7-10, 15 (2020). A practice which has raised some concerns is price discrimination. Interestingly, though, while price discrimination of that sort is used by suppliers to maximize their profits, it could also be advantageous to certain users who gain access to products and services that would otherwise be offered to them in prices they cannot afford.

92. See Omri Ben-Shahar & Ariel Porat, *Personalizing Mandatory Rules in Contract Law*, 86 U. CHI. L. REV. 255, 274-76 (2019) (discussing how information on individuals' needs and traits could lead to more desirable, personalized contract provisions). Note that the introduction of personalized law may carry different implications depending on the context in which it is implemented. For instance, the use of data for personalized law in the context of private ordering (e.g., completing incomplete contracts) may align with the interests of the contracting parties seeking to maximize information for their risk allocation. On the other hand, applying personalized law in cases involving negligence or criminal liability may invoke disparity of interests and thereby raise different policy considerations.

93. See, e.g., Porat & Jacob Strahilevitz, *supra* note 25; Ben-Shahar & Porat, *supra* note 25.

utilize fine-grained data on individuals to develop and apply personally tailored legal norms.⁹⁴

For instance, real-time data-collection on individuals' habits, such as driving patterns, could be used to craft personalized speeding limits for each driver, derived from their driving abilities as well as risks they pose.⁹⁵ Rather than setting a single speed limit that applies to all drivers, speed limits might be personally tailored to individual road-users, based on their experience, driving history, or real-time road conditions. Privileged parking permits, as currently offered to disabled persons, could be issued to individuals based on relevant temporary or permanent health conditions or family circumstances (e.g., driving young children). Such tailored norms could be embedded in the digital infrastructure (e.g., autonomous cars, smart parking facilities, and roads) and individually applied in real time by enabling parking, issuing a ticket, or even slowing down a speeding car, for instance.⁹⁶ Although not without challenges and concerns,⁹⁷ personalized laws may also increase efficiency in law enforcement by reducing the under- or over-inclusive risk-avoidance mechanisms and reducing institutionalized discrimination.⁹⁸

Personalization capabilities take customization a step further. Consider, for instance, Reidenberg's *Lex Informatica*, according to which customizations afforded individuals the flexibility to decide whether to be subject to a certain

94. See Ben-Shahar & Porat, *supra* note 25, at 634–36.

95. See *id.* at 630–31.

96. See, e.g., Anthony J. Casey & Anthony Niblett, *The Death of Rules and Standards*, 92 IND. L. J. 1401, 1401 (2015) (discussing the concept of “microdirective,” where machines will provide individuals with instructions on how to comply with the law in a tailored manner which factors the specific circumstance and context). For instance, a sensor in a car will register data and send it over the internet to an algorithm, which will analyze the information, combine it with other sources of relevant data, and determine whether the driver violated legal limits. A digital system could then potentially issue a fine or even remotely disable a car after issuing a warning.

97. Personalized law may undermine important values, raising considerable concerns regarding privacy, equality under the law, and civil liberties. See, e.g., SHOSHANA ZUBOFF, *THE AGE OF SURVEILLANCE CAPITALISM* 238 (2019); Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias*, PROPUBLICA (May 23, 2016), <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>; *Big Data: A Tool for Inclusion or Exclusion* (Jan. 6, 2016), <https://www.ftc.gov/reports/big-data-tool-inclusion-or-exclusion-understanding-issues-ftc-report>; Karen Li Xan Wong & Amy Shields Dobson, *We're Just Data: Exploring China's Social Credit System in Relation to Digital Platform Ratings Cultures in Westernised Democracies*, 4 GLOB. MEDIA & CHINA 220 (2019); DANIEL SOLOVE, *NOTHING TO HIDE: THE FALSE TRADEOFF BETWEEN PRIVACY AND SECURITY* (2011).

98. Porat & Strahilevitz, *supra* note 25; Ben-Shahar & Porat, *supra* note 25; see also Jane Bambauer, *Other People's Papers*, 94 TEX. L. REV. 205 (2015). For the purpose of this article, we do not challenge the assumption that personalized law can make legal enforcement more efficient or question the desirability of increased and more efficient law enforcement.

norm (e.g., blocked violent content) and to shape the norm according to their specific requirements. For example, parents of young children could apply a screening mechanism to block R-rated or PG-13 content and also deactivate screening at a particular time according to their children's usual bedtime. Parents of older children could choose to block R-rated content only and enable viewing blocked content later at night when the children are asleep. Adults with no small children could ignore the blocking system altogether.

Taking advantage of personalization, Lex AI provides an opportunity to tailor norms and outcomes more precisely to each individual's profile. In its decision-making process, for example, Lex AI may consider information that individuals were unaware of, did not remember, or did not realize was important. For instance, an AI medical diagnosis application may account for data it measured on a patient's quality of sleep. Similarly, if the application had access to the patient's medical history, it could note the fact that when the patient was a child their parents had reported that they were allergic to a certain substance—a fact that the patient never knew or perhaps forgot. Moreover, because Lex AI “knows” relevant data on other individuals, in aggregating the data, it may reach conclusions that are highly relevant to the individual in question. For example, an AI application may identify, based on the experiences of other users, that people with a particular occupation suffering from a certain condition are at greater risk if treated with a specific medication.

In sum, Lex AI presents the opportunity to personalize choices tailored to each individual—in both the private and public spheres. Would it therefore be accurate to argue that, like its predecessor Lex Informatica, Lex AI is a manifestation of private ordering enabling more legitimate and efficient outcomes by reaching the most informed decisions about each individual? As we demonstrate in the next Section, such a conclusion should be taken with a grain of salt.

In theory, private ordering is more likely to maximize the efficiency and legitimacy of norms, provided that it reflects the individual's choice.⁹⁹ This theory rests on several assumptions: first, the individual has substantial knowledge of the mechanisms or solutions subject to the individual's choice; second, the individual may personally choose from among these options; and third, the individual's choice expresses their true preferences. We argue that Lex AI, as currently deployed, structures the decision-making process in a manner that renders these three assumptions debatable.

99. Notably, and as further discussed in Section III.C.2, an individual's choices may not always reflect their best interests or the best way to achieve their preferences.

2. *Does AI Enable Informed Decision-Making?*

The theory that private ordering facilitates efficient and legitimate norms further assumes that, when making their decisions, individuals have sufficient information on potential alternatives and can thus base their choice on more than guesswork or pure luck. In other words, this theory assumes that, armed with information on the different options, individuals are able to estimate the potential outcomes or implications of each alternative and reach a decision that maximizes their interests.¹⁰⁰

Granted, parties to private ordering often lack *full* information on all the relevant aspects pertaining to the norms they shape or the commitments they undertake. In fact, the very essence of contracts is about allocating risk among the parties in the face of unforeseeable circumstances.¹⁰¹ Yet, even if the exact outcome of the contract is unknown at the time of signing it, the parties know what the outcome would look like given a certain set of circumstances. For example, although the parties cannot know if a hurricane will hit the construction site that is the subject of the contract, they know what the consequences of choosing a certain contract provision over another would be if a hurricane were indeed to strike.

Similarly, the “if-then” structure of Lex Informatica also enabled individuals to make informed choices based on full or partial information about options and a view of the options’ consequences. For example, a PICS user choosing to filter violent content could conceive in advance what qualified as violent content based on predetermined labeling criteria (e.g., violent scenes involving excessive blood). The PICS user might then conceive that *if* they chose to filter violent content, *then* the PICS algorithm would block content depicting bloody scenes. In other words, when choosing whether to filter certain content, the individual had a clear vision of what their choice would entail. Moreover, choices made at a certain moment in time are generally likely to continue yielding the same results in a subsequent future, at least until a change in the design of the system occurred. Violent content, for example, would remain classified as such until the entity in charge of the classification made an active decision to change it.

100. Elkin-Koren, *supra* note 42, at 1181. (“Also, from a political perspective, if users are not aware of the implications of their choices, their choices cannot reflect the exercise of an autonomous will.”).

101. See Deepankar Sharma & Priya Bhatnagar, Risk Allocation & Subsequent Legal Issues in Construction Contracts (Feb. 28, 2014) (unpublished manuscript), <https://ssrn.com/abstract=2403045>.

Under the principles of Lex AI, by contrast, the individual's knowledge of potential alternatives and their ability to rely on those alternatives continued applicability are very limited. To illustrate this point, consider a scenario in which, as with contracts and Lex Informatica, the individual is *aware of* and *involved in* setting the norms.¹⁰² Under a contract, a party can choose to block violent content and define "violent" as they please.¹⁰³ Under Lex Informatica, a party can use technological means to do the same by relying on pre-determined classifications of "violent" content. In Lex AI, however, a user of an AI-based filtering system that seeks to block violent content would not be able to predict in advance what the end result of their decision to block violent content would be. This is due to several reasons. First, users may not realize in real time that the system does not necessarily serve their own best interests—whether because it promotes the system operators' own interests¹⁰⁴ or because of an explore-and-exploit trade-off typical of learning models.¹⁰⁵ Further, even if the system's goals are well-known, users might still lack information on expected outcomes: although the science of prediction and probabilities was introduced as a way to increase certainty—including legal certainty¹⁰⁶—AI outcomes cannot, in fact, be accurately predicted, even by their own creators,

102. The level of user engagement in AI decision-making can be conceptualized as a spectrum. On one side are those systems where the uses thereof are not only involuntary but also do not require active input by the user (e.g., systems such as COMPAS that assist judges in criminal cases). Next, there are AI systems that the user may choose whether to use or not but have no active effect on their selections (e.g., Netflix, where users are not asked for an active input on their desires). Waze is an example of a system closer to the other side of the spectrum, where users can not only choose whether to use the system in the first place but can also convey their desires in a manner that would affect the system's consequences (e.g., by requesting that Waze direct them via the shortest rather than the quickest route or direct them through a scenic rather than the shortest route).

103. Granted, a failure to properly define terms would result in leaving such definitions to the interpretation and discretion of a judge. Under such a scenario, the norm would no longer be shaped by the parties but rather by the court in a manner involving both *ex ante* and *ex post* input. But, if the provisions are carefully drafted, parties to a contract have full or at least much information on what the agreement into which they have entered entails.

104. See Michal S. Gal, *Algorithmic Challenges to Autonomous Choice*, 25 MICH. TELECOMM. & TECH. L. REV. 59, 83 (2018).

105. See *infra* notes 135, 162–163 and accompanying text.

106. Probabilistic tools may assist in assessing the likelihood of potential occurrences or behaviors, or for a more specific example, in assessing the impact of sanctions. See Mike Ananny, *Probably Speech, Maybe Free: Toward a Probabilistic Understanding of Online Expression and Platform Governance*, KNIGHT FIRST AMENDMENT INST. (Aug. 21, 2019), <https://knightcolumbia.org/content/probably-speech-maybe-free-toward-a-probabilistic-understanding-of-online-expression-and-platform-governance>.

let alone by users.¹⁰⁷ One explanation lies in humans' limited cognitive ability to process the magnitude of information processed by AI.¹⁰⁸

Another explanation is that users not only lack information on the system's expected decision, they also lack information on the data the system uses to reach such decisions.¹⁰⁹ Users do not necessarily know what type of data the system collects, which sources it collects from, and what specific information is collected in each and every case.¹¹⁰ The more limited the information available to users on the data fed into the system or inferred by the system, the less able they are to estimate the results yielded by the system under any of their potential choices.¹¹¹

Moreover, the information that the system uses to make its decisions is constantly changing. Content once classified as "violent" could subsequently be classified as "nonviolent," or vice versa, based on new or transformed

107. See Gal, *supra* note 104, at 63; see Karni Chagal-Feferkorn, *The Reasonable Algorithm*, 1 U. ILL. J.L. TECH. & POL'Y 111, 133–35 (2018).

108. See *Artificial Intelligence Singles Out Neurons Faster Than a Human Can*, SCI. DAILY (Apr. 12, 2019), <https://www.sciencedaily.com/releases/2019/04/190412150628.htm> (describing an algorithm that can identify and segment neurons in minutes, a task that could take a trained human researcher twenty-four hours of nonstop work). For a general discussion of the superior abilities of AI systems compared to humans in the decision-making process, see Karni Chagal-Feferkorn, *How Can I Tell If My Algorithm Was Reasonable*, 27 MICH. TECH. L. REV. 213 (2021).

109. See Carey Shenkman, Dhanaraj Thakur & Emma Llansó, *Do You See What I See?: Capabilities and Limits of Automated Multimedia Content Analysis*, CTR. FOR DEMOCRACY & TECH. (May 20, 2021), <https://cdt.org/insights/do-you-see-what-i-see-capabilities-and-limits-of-automated-multimedia-content-analysis/> (arguing that despite research efforts to promote the comprehensibility of ML tools, the technical steps describing how ML systems make decisions or weigh various features may involve billions of interrelated parameters of which humans cannot conceive).

110. To take the example of content moderation, even if a user who wishes to upload a video to YouTube knows the system is searching for similarities between her video and a dataset of protected works, she would not necessarily know what kind of information is relevant to the system. The system could be looking for information pertaining only to the work itself, whether profiles of others who are similar or related to the user as friends or family members have uploaded infringing materials in the past, or, if the system also collects information on the uploader's past behavior, prior incidents of uploads with suspected similarities to protected works. The user would also not necessarily know the source of information—whether, for example, prior incidents that may indicate a past of copyright infringement are considered only based on data from YouTube alone or from other platforms as well. Moreover, even if aware of the type and sources of information relevant to the system's decision, the user will likely not know the specific content harvested from these various sources.

111. Moreover, users lack information not only on the data collected but also on the weight assigned by the system to any particular piece of data.

information that may not pertain to the content itself.¹¹² For example, the system may account for new types of the user's (or other users') habits, or it could consider recent literature on the harmful effects of online content when classifying what content should fall under the classification of "violent"—all in a manner not foreseeable to the user.

Lastly, AI decision-making's feedback loop mechanism, through which its previous decisions are fed back into its decision-making process, renders its outcomes less foreseeable and may thus potentially influence subsequent decisions.¹¹³ In fact, AI decision-making can generate and shape norms all on its own, rendering them even less foreseeable.¹¹⁴

Lex AI decisions may not announce an explicit norm *ex ante* (e.g., that any content depicting a weapon is forbidden), but it may instead seek to optimize a certain objective function (e.g., by removing all terrorism-related content based on known hashes¹¹⁵). The norm itself (e.g., the definition of illicit materials) may change depending on the system's learning—for instance, whether it has practiced the removal of similar materials. This may result in rapidly and independently evolving norms.

Consider, for instance, personal-health applications that involve tracking behavior, predicting a certain health risk, and simultaneously regulating the behavior to reduce that risk. As insurance companies determine premiums based on risk, apps that monitor driving patterns may help reduce auto

112. See Katharine Trendacosta, *Unfiltered: How YouTube's Content ID Discourages Fair Use and Dictates What We See Online*, ELEC. FRONTIER FOUND. (Dec. 10, 2020), <https://www.eff.org/wp/unfiltered-how-youtubes-content-id-discourages-fair-use-and-dictates-what-we-see-online>.

113. See Masoud Mansoury, Human Abdollahpouri, Mykola Pechenizkiy, Bamshad Mobasher & Robin Burke, *Feedback Loop and Bias Amplification in Recommender Systems*, 29 PROCS. ACM INT'L CONF. ON INFO. & KNOWLEDGE MGMT. 2145 (2020), <https://arxiv.org/pdf/2007.13019.pdf>; see also *infra* notes 138, 162–163 and accompanying text.

114. See Elkin-Koren & Perel, *supra* note 84 (“Note that governing speech by AI does not merely apply existing norms, thereby simply reflecting existing values and trade-offs. AI systems also generate new norms, which are implicit in the speech affordances generated by the system. Once content is tagged as hate speech or terrorist propaganda, it may not only become unavailable but also be considered illegitimate. Thus, AI systems which seek to identify hate speech, may also carry a regulatory consequence: shaping users' behavior by distinguishing between legitimate and illegitimate expression (where only the latter are removed).”).

115. Hashing is the practice of translating a potentially long input into a short hash value. Claudio Buttice, *Hashing*, TECHOPEDIA (Apr. 27, 2021), <https://www.techopedia.com/definition/14316/hashing-cybersecurity>. For example, hashing could constitute translating twenty seconds of an R-rated film to the hashes “#violence,” “#foul language,” and other hashes.

insurance costs, just as apps designed to track sports or fitness activity—commonly available on devices like Fitbit or Apple Watch—may help reduce life insurance costs.¹¹⁶ Assume, for example, that the insurer offers a premium reduction to those who exercise at the highest level—say, the top ten percent. As more members take up regular exercise, the level of activity required from each member to reach the top ten percent rises. Based on the reasons detailed above, the user's choices that instruct a Lex AI are not only based on very little information to begin with but they also result in ever-changing outputs that even the system's creators cannot predict. Unlike parties to a contract defining what content shall be blocked or Lex Informatica users choosing what content to block based on predefined labels, users under Lex AI will often not be able to predict the outcome of their choices.¹¹⁷ That is, will a certain piece of content be blocked or not if a choice is made to filter violent content? Users under Lex AI are making less-informed choices and cannot know what in fact those choices would entail.

3. *Does Lex AI Even Let Users Choose?*

Not only are choices under Lex AI generally less informed than those pertaining to other types of private ordering, but even the very ability to choose is impaired.¹¹⁸

116. Regulation normally set clear norms and design incentives linked to compliance. It regulates behavior by applying negative consequences (e.g., sanctions) for non-compliance, and sometimes positive incentives to encourage compliance (e.g., bonuses, safe harbor, tax cuts, savings in interest rates). When regulation is attached to behavior, incentives would apply ex post but encourage rational players to shape their behavior accordingly. Lex AI may be designed to predict behavior which has never occurred, thereby facilitating the prevention of behavior before it is executed. Sanctions and awards may be distributed based on the potential that a particular behavior will take place. Examples include a preventative arrest of a suspicious tourist at the airport, the recruitment or promotion of an employee based on their past performance, or performance in other spheres.

117. *See supra* notes 103–114 and accompanying text.

118. A preliminary question is that of the choice of whether to be subject to the system in the first place. This dimension does not concern the manner in which the AI decision-making process operates, and it is therefore only anecdotal for our discussion. Often, however, this element of choice is lacking with respect to AI systems. This is because the use of various AI systems is not left to individual discretion. For example, COMPAS is deployed in the United States' criminal justice system without the individual's voluntary agreement to be subject to the system's decision-making. Furthermore, even in contexts where the individual does have the choice *not* to be subject to AI decision-making, these systems are often used to make decisions of high importance—for example, in the context of housing, credit, health, or human resources. In these important contexts, refusing to be subject to the AI system could be to the individual's detriment. Consider, for example, a user applying for a job in a market where jobs are scarce, but the hiring process is only enabled through algorithmic decision-

“Choice” involves the ability to choose between different alternatives in a manner that would affect the outcome. For example, a user choosing to filter “violent,” but not “obscene,” content is theoretically making their own choices.

Arguably, Lex AI weakens the ability of users to make such autonomous choices. This is due to the particular way Lex AI creates what Richard Thaler and Cass Sunstein call a “choice architecture.”¹¹⁹ Drawing on studies in cognitive psychology and behavioral economics, Thaler and Sunstein highlighted the surrounding context’s significance for shaping people’s choices. They argued that, by altering a choice architecture, policymakers could overcome citizens’ bounded rationality, “nudging” them towards the “right” decision in a noncoercive way. “Nudges,” according to Sunstein, are “interventions that steer people in particular directions but that also allow them to go their own way.”¹²⁰

Advocates of nudging argue that it is justifiable for the purpose of improving public welfare. Its appeal lies in the fact that it ostensibly involves no coercion, preserving freedom of choice.¹²¹ Yet, although nudging—algorithmically or otherwise—is not coercive, it undoubtedly affects human choices, whether by shaping individual preference or by prompting people to choose the alternatives that presumably best reflect their preferences.¹²² When users are presented with certain alternatives and not others, they are evidently more likely to choose from among the ones presented. For instance, the entire Search Engine Optimization (SEO) industry revolves around the fact that users are likely to select the options presented to them first.¹²³

making. Having chosen to subject herself to the AI process, it could be argued that any algorithmic decision does not reflect her true choice.

119. RICHARD THALER & CASS SUNSTEIN, *NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS* 6, 81 (2008).

120. Cass Sunstein, *The Ethics of Nudging*, 32 *YALE J. REGUL.* 413, 417 (2015) (“To qualify as a nudge, an intervention must not impose significant material incentives (including disincentives). A subsidy is not a nudge; a tax is not a nudge; a fine or a jail sentence is not a nudge. To count as such, a nudge must fully preserve freedom of choice. If an intervention imposes significant material costs on choosers, it might of course be justified, but it is not a nudge.”).

121. See generally Ayala Arad & Ariel Rubinstein, *The People’s Perspective on Libertarian-Paternalistic Policies*, 61 *J. L. & ECON.* 311 (2018).

122. See Anne van Aaken, *Judge the Nudge: In Search of the Legal Limits of Paternalistic Nudging in the EU* 1620 (Univ. of St. Gallen L. Sch., Working Paper No. 2015-01, 2015), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2563296.

123. See, e.g., Madhu Bala & Deepak Verma, *A Critical Review of Digital Marketing*, 8 *INT’L J. MGMT., IT & ENG’G ENG’G* 321, 329 (2018) (reviewing current marketing trends and related consumer motives, including the use of SEO).

Arguably, the output of AI systems presents users with different levels of choice.¹²⁴ At one end of the spectrum, some AI systems may reach decisions and execute them on behalf of the user without requesting their approval or input.¹²⁵ At the other end of the spectrum are systems that present all possible options to the user and prompt them to make their own choice—including, for example, Google search.¹²⁶ Located in the middle of the spectrum are systems that rank some of the alternatives and let the user choose among them. Lex AI influences users' choices, steering them in directions it deems preferable whether the system presents but one option and asks the user to approve it, narrows the list of alternatives to several options to choose from, or merely selects the order in which all alternatives appear.¹²⁷

Moreover, the ability to tailor choice selections to particular individuals based on their past behavior renders individuals more vulnerable to influence. Karen Yeung has argued that ML recommendation systems—what she calls decision-guidance processes¹²⁸—enable a “hypernudge,” which shapes individual decision-making to serve the interests of the hypernudge’s operators. In that sense, AI systems regulate behavior by circumventing rational choice and using mechanisms to influence people’s behavior behind their backs. Indeed, studies show that users might be less capable of deviating

124. Users of Waze, for example, are both free not to use the app in the first place and free to ignore its directions. In that sense, Lex AI is often a mere “nudge” that respects freedom of choice and simply incentivizes users to follow its suggestions. Daniel M. Hausman & Brynn Welch, *Debate: To Nudge or Not to Nudge*, 18 J. POL. PHIL. 123, 135 (2010); Luc Bovens, *The Ethics of Nudge*, in PREFERENCE CHANGE 207, 216 (Till Grüne-Yanoff & Sven Ove Hansson eds., 2009); Riccardo Rebonato, *A Critical Assessment of Libertarian Paternalism*, 37 J. CONSUMER POL'Y 357 (2014); Jeremy Waldron, *It's All for Your Own Good*, THE N.Y. REV. (Oct. 9, 2014), <https://www.nybooks.com/articles/2014/10/09/cass-sunstein-its-all-your-own-good/>.

125. See Gal, *supra* note 104, at 69.

126. See *id.*

127. Users' choices might be shaped by the alternatives presented to the user as opposed to the ones the system omits, as users tend to limit their choice only to the visible alternatives. The order of the options presented too can stir users' choices in a certain direction, as users tend to choose among the alternatives presented first. See Michael R. Baye, Babur De los Santos & Matthijs R. Wildenbeest, *Search Engine Optimization: What Drives Organic Traffic to Retail Sites?*, 25 J. ECON. & MGMT. STRATEGY 6 (2016) (quantifying the effect of a search result rank on the number of clicks).

128. See Karen Yeung, “Hypernudge”: *Big Data as a Mode of Regulation by Design*, 20 INFO., COMM'N & SOC'Y 118, 121 (2017). Yeung distinguishes between automated decision-making and decision-guidance processes that “seek to *direct or guide* the individual’s decision-making processes in ways identified by the underlying software algorithm as ‘optimal’, by offering ‘suggestions’ intended to prompt the user to make decisions preferred by the choice architect . . .” See *id.*

from recommendations made by ML recommendation systems.¹²⁹ Consequently, even if the decision is made by the user and not by the system, it does not necessarily reflect a neutral, uninfluenced choice.¹³⁰

4. *Do Lex AI's Choices Reflect Users' Preferences?*

Despite the limited information available to users to facilitate informed choice and the limited nature of the choice itself, one could argue that, as long as the system—through its advanced personalization capabilities—executes users' true preferences, then the efficiency and perhaps the legitimacy obtained through private ordering will be achieved.¹³¹

Yet, the assumption that personalized decisions Lex AI make reflect users' *true preferences* is dubious. First, Lex AI is deployed by various stakeholders motivated by various interests, including commercial and political interests. Even among those systems that purport to reflect users' preferences as their primary objective may, in practice, maximize other goals. Consider, for example, Scatter Lab's Science of Love app, which was purportedly designed to promote individuals' preferences by educating them on the level of affection their partner feels for them based on AI analyses of the text messages exchanged between couples. In fact, the goal of the system was to scrape private text messages for use in ML training.¹³²

Second, even assuming there exist AI systems whose sole goal is to maximize users' preferences,¹³³ there are several reasons why this goal may not

129. For instance, a recent study showed that consumers tend to favor an algorithm's advice over human experts, even after poor performance of their investment. See Packin, *supra* note 24, at 344.

130. Recent studies on the nudge theory show that, although free choice is theoretically possible when encountering a nudge, the human brain is in many cases simply unable to process the possibilities and consequences associated with the options not "nudged" toward, in fact leaving the user no free choice. See Thaler & Sunstein, *supra* note 119, at 94–96.

131. Imagine, for example, that a person finds and takes possession of another person's private diary, where all the wishes of the author are described. If the finder of the diary executes the writer's wishes, then—at least in terms of efficiency and provided that the author's preference has not changed over time—their execution would enjoy the same efficiency as if executed by the writer herself. Since the writer did not make a choice to execute her preferences, it of course does not reflect the autonomy and free choice celebrated in private ordering. It could be argued, however, that subjecting a person to her own preferences is more legitimate than subjecting her to norms that do not reflect her preferences.

132. See Katyanna Quach, *Science of Love App Turns to Hate for Machine-Learning Startup in Korea After Careless Whispers of Data*, THE REGISTER (Feb. 15, 2021), https://www.theregister.com/2021/02/15/in_brief_ai/.

133. Note that such systems would still need to deviate from maximizing preferences for the sake of learning and improving, as discussed below in Section III.C.

be achievable. For example, the limitations of algorithmic capabilities mean that, presently, AI systems lack the ability to understand many human nuances or make decisions that require intuition or empathy.¹³⁴ A dating app, for example, may accurately analyze a user's preferences for certain looks based on geometrical symmetries or other quantifiable parameters, but it will not necessarily recognize "fuzzy" traits such as charisma or other nonconcrete features that will result in a human "click."

Other reasons for AI systems' failure to maximize users preferences stem from the algorithmic decision-making process itself. After all, a single user's preference is but a prediction. AI systems do not presume to know what a single individual truly wishes. Rather, they use information on the individual's prior behavior and choices, coupled with information on what other users did or chose in the past, to *estimate* what the individual's current preference.¹³⁵ It has been argued, for instance, that "Facebook is not giving the user what the user wants—Facebook is giving the user what it thinks a demographic stereotype wants."¹³⁶

134. Referring to human decision-makers and contrasting them with AI systems, Professor Joshua Davis writes, "We neither know precisely what we are measuring when we say we want to maximize pleasure nor do we know how to measure it. Of course, human beings too must face these difficulties. But, in doing so, we have resources that UAI does not. We can assess what objectives are worth pursuing. We have intuitions about what is just or fair in particular circumstances. We can empathize with other living beings, imagining what their experiences might be and how they might rank them. We rely on these and other capacities to come to moral conclusions, however imperfect." Joshua P. Davis, *AI, Ethics, and Law: A Possible Way Forward*, in *ARTIFICIAL INTELLIGENCE AND PRIVATE LAW: GLOBAL PERSPECTIVES* (Larry Di Matteo ed., forthcoming 2022); *see also* Michal S. Gal & Niva Elkin-Koren, *Algorithmic Consumers*, 30 *HARV. J. L. & TECH.* 309, 323 (2017) ("The algorithmic choice may not always accurately reflect consumers' preferences. . . . One reason is inherent limitations of computer coding. For instance, algorithms might not (as of yet) be able to recognize and relate to certain nuances that humans intuitively understand. While such nuances might not be important in many transactions, they could be essential in others. Accordingly, most of us would probably not want an algorithm to automatically choose our partner in business or in life, and possibly not our wedding ring.").

135. If, for example, a user's viewing history, demographic parameters, and social interactions are similar to a group of other users, and many of these users have downloaded a new TV series and rated it highly, then the AI system will likely assume that the underlying user would also like the show and recommend it to her. The problem is that the system only *predicts* that the underlying user, like her peers, would enjoy the series. The prediction is not in any way based on the user's *true* preferences, only on *presumed* ones, which are deducted from the preferences of others.

136. Don Owens, *Facebook Engages in Online Segregation and Redlining Through Discriminatory Advertising System, Lawyers' Committee Argues*, *LAWYERS' COMM. FOR C.R. UNDER L.* (July 10, 2020), <https://www.lawyerscommittee.org/lawyers-committee-confronts-facebooks-attempts-to-dismiss-digital-redlining-lawsuit-against-its-housing-advertisements/> (quoting

Reliance on past choices ignores the dynamic nature of choosing and might perpetuate preferences that are no longer valid. This is especially true given AI systems' path-dependency (further discussed in Section III.C), whereby an algorithm's recursive feedback loop might perpetuate outdated choices and lead to results that no longer reflect the user's preference, if they ever did.¹³⁷ Relatedly, a user's preference to try something new—which does not correspond to their previous patterns of choices—will likely not be picked up by the system.¹³⁸

The net result is that, even if accurate predictions can be made by drawing on an individual's preferences to indicate those of another—which is not always the case—then the choices or outcomes the system generates for an individual are constantly subject to change based on others' preferences. A user wishing to watch a comedy, for example, will be offered a variety of comedies based on other users' preferences that the system deems similar to that user. If multiple users decide to block a certain comedy for reasons that could be entirely unrelated to our user's preferences, then the system will stop offering that comedy to our user, even though *their* preferences have not changed, only those of others.

Coupled with the “nudging” effect of algorithms discussed in Section III.B.3, which might cause the user to not actively look for content that the system does not offer them, Lex AI thus might steer individuals toward choosing among alternatives that reflect the preferences of others and not their own; and this occurs in a manner that could be perpetuated by AI systems' path-dependent nature. In other words, Lex AI ought not be mistaken as necessarily reflecting users' preferences. Rather, as further discussed below, in many cases it may be *shaping* those preferences.

David Brody, counsel and senior fellow for privacy and technology at the Lawyers' Committee). Brody argued further that “[r]edlining is discriminatory and unjust whether it takes place online or offline and we must not allow corporations to blame technology for harmful decisions made by CEOs.” *Id.*

137. To take an intuitive example, consider a Netflix viewer who watched a horror movie despite disliking the genre. If, on their next viewing, the user is presented with a selection of “movies you might like,” all of which are horror movies, the user may be tempted to select one of the options, despite their distaste—whether due to laziness, curiosity, etc. In that sense, Lex AI shapes norms and behaviors. But, since the process of Lex AI is recursive, the user's additional selection of a horror movie will render the system even more confident that it correctly classified the user as a horror-movie enthusiast. The system will then continue to offer these types of movies when the viewer next returns to Netflix.

138. *See, e.g.,* Owens, *supra* note 136 (“Facebook profiles its users on the basis of their protected characteristics and then provides different services to these users and excludes them from economic opportunities, like financial services, based on those same characteristics.”).

C. LEX AI AS A CENTRALIZED GOVERNANCE MODEL

1. *Beyond Individual Choice*

Rulemaking processes under private ordering are decentralized by their very nature as they are grounded in the exercise of choice among those who craft the norms. As demonstrated in Section III.B, Lex AI does not accurately reflect users' choices. Instead, it reflects the outcome of a centralized decision-making process.

Indeed, Lex AI is often deployed to coordinate different social actors' behavior, as with recommendation systems or the management of traffic, health, or law enforcement in smart cities.¹³⁹ Such coordination may involve not only collecting information on people's preferences, but also determining trade-offs between actors' conflicting interests and giving priority to some over others.

In *Considerations on Representative Government*, philosopher John Stuart Mill provocatively questioned the desirability of a "good despot"—an "all-seeing" ruler, exercising "superhuman mental activity," and fully informed of the entire affairs of a "mentally passive people."¹⁴⁰ Although he emphasized the advantages of a good despot, Mill was skeptical of the feasibility of such a figure, suggesting that good despots would in practice be unlikely to consider all affected interests.¹⁴¹ According to Mill, each individual is the "only safe guardian of his own rights and interests," and therefore, once excluded from decisions pertaining to her interests, these interests might be overlooked. Consequently, supreme control over power should be "vested in the entire aggregate of the community."¹⁴²

139. See, e.g., Tibi Puiu, *AI Traffic Management Could Finally Declog Urban Roads*, ZME SCI. (Feb. 26, 2021), <https://www.zmescience.com/science/news-science/ai-traffic-management-could-finally-declog-urban-roads/>; Da-Young Kang, Kyung-Jae Cho, Oyeon Kwon, Joon-Myoung Kwon, Ki-Hyun Jeon, Hyunho Park, Yeha Lee, Jinsik Park & Byung-Hee Oh, *Artificial Intelligence Algorithm to Predict the Need for Critical Care in Prehospital Emergency Medical Services*, 28 SCANDINAVIAN J. TRAUMA RESUSCITATION & EMERGENCY MED. 17 (2020), <https://sjtrem.biomedcentral.com/articles/10.1186/s13049-020-0713-4>; CHRISTOPHER KIRWAN & FU ZHIYONG, SMART CITIES AND ARTIFICIAL INTELLIGENCE: CONVERGENT SYSTEMS FOR PLANNING, DESIGN, AND OPERATIONS (2020).

140. Mill, *supra* note 27, at 46 ("He must be at all times informed correctly, in considerable detail, of the conduct and working of every branch of administration, in every district of the country, and must be able, in the twenty-four hours per day, which are all that is granted to a king as to the humblest laborer, to give an effective share of attention and superintendence to all parts of this vast field What should we then have? One man of superhuman mental activity managing the entire affairs of a mentally passive people.").

141. *Id.*

142. *Id.*

Arguably, Lex AI could bridge this gap. It introduces a superior form of collective governance because it bases its decision on the efficient collection and analysis of granular information regarding actual preferences and behavior. As such, it could possibly address one of the major challenges associated with centralized governance: information failure due to limited and outdated information on the actions and appetites of each individual.¹⁴³ Coupled with more robust data collection and data analytics capabilities, Lex AI can enable the calculated balancing of various interests under complex scenarios. Lex AI thus offers a mechanism for coordinating the behavior of social actors, which is superior, at least in certain respects, to democratic government.¹⁴⁴ Indeed, Lex AI advocates point to its potential to reduce “the cost of core governance functions, improve the quality of decisions, and unleash the power of administrative data, thereby making government performance more efficient and effective.”¹⁴⁵ Arguably, Lex AI could more effectively reach decisions based on a collective good, in a manner that takes more relevant data into account. And, by factoring in information on each individual, it could offer de facto participation in shaping norms.

Moreover, by utilizing fine-grained data on individuals’ behavior, Lex AI can potentially develop and apply shared or personally tailored norms more efficiently.¹⁴⁶ Like other personally tailored treatments (e.g., medical care, nutrition, etc.), AI capabilities may personalize collective action such that each individual would be subject to different norms or rules based on their personal information. Preferences, capabilities, past behavior, or other parameters will be accounted for by the system in a manner that could be more desirable to individuals and enhance efficiency.¹⁴⁷

In this way, Lex AI could be viewed as a centralized system of governance that facilitates decentralized coordination of social actors. But what would be the guiding principles of such coordination? Coordination by markets, for instance, presumably signals consumers’ preferences by price as determined

143. See *supra* note 26.

144. Kevin Werbach, *Panopticon Reborn: Social Credit as Regulation for the Algorithmic Age*, 2022 UNIV. ILL. L. REV. (forthcoming 2022).

145. ENGSTROM ET AL., *supra* note 75. In this context, a more nuanced discussion is necessary. Karen Yeung, for instance, highlights the advantages of outcome-based public policy and data-driven public management. See Karen Yeung, *Algorithmic Regulation: A Critical Interrogation*, 12 REGUL. & GOVERNANCE 505, 509–11 (2017); Kenneth A. Bamberger, *Technologies of Compliance: Risk and Regulation in a Digital Age*, 88 TEX. L. REV. 669 (2010).

146. See Porat & Strahilevitz, *supra* note 25, at 1440–41; Ben-Shahar & Porat, *supra* note 25 at 679–80.

147. See Porat & Strahilevitz, *supra* note 25, at 1450–53.

through the market mechanism of supply and demand, and politics in liberal democracies ostensibly reflects the aggregated will of the governed through voting.

The signaling of preferences under Lex AI, however, is very limited in nature. First, even when the system does not make selections itself but leaves room for users' input, the nudging effect calls into question whether users' input truly reflects their own preferences. Second, as elaborated below, the choices offered to users are path dependent, perpetuating past choices regardless of the users' current preferences.¹⁴⁸ Thus, users' choices under Lex AI provide an unreliable signal for how users would have wanted norms to be shaped, how conflicting interests ought to be resolved, and how different values are to be prioritized.

If Lex AI does not necessarily reflect users' choices while also shaping users' behavior, what goals does the system promote and how are those goals being determined? As we elaborate next, we believe the way Lex AI determines individuals' best interests and the common good raises critical concerns about the decision-making method—one that is (a) path-dependent and based on an optimization function, and (b) fails to provide space for deliberating upon trade-offs.

2. *How Does Lex AI Decide Users' Best Interests?*

In the absence of real choice manifested through Lex AI, how does Lex AI determine what is best for any individual user? And, more generally, when used to coordinate behavior, how does it determine what is best for all users?

One obvious answer is that Lex AI is designed to serve the goals and interests of those who deploy it. Indeed, many commentators, scholars, and policymakers have warned that governments could use AI systems to perform governmental tasks in opaque or unfair manners.¹⁴⁹ Despite its apparent potential advantages, Lex AI governance raises serious concerns regarding privacy, equality under the law, and civil liberties.¹⁵⁰

Others have warned that private actors, especially tech giants such as Google or Facebook, are basically governing human behavior with AI without

148. See *infra* notes 135, 162–163 and accompanying text.

149. See, e.g., Danielle Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249 (2008); PASQUALE, *supra* note 19; Carrie B. Sanders & James Sheptycki, *Policing, Crime and 'Big Data': Towards a Critique of the Moral Economy of Stochastic Governance*, 68 CRIME, L. & SOC. CHANGE 1, 7–9 (2017).

150. See *supra* note 97.

any accountability.¹⁵¹ For instance, digital platforms operating in multi-sided markets profit from selling user attention and user data to advertisers.¹⁵² AI content moderation thus aims to attract more users to the platform and keep existing users engaged for longer periods of time in order to generate more revenues for the platform.¹⁵³

Yet, putting aside the practical question of the interests and motivations of the entities governing through Lex AI, AI *could* theoretically be configured to steer behavior toward what is perceived to be in users' best interest.¹⁵⁴ In fact, given AI's unique features, Lex AI arguably makes choices for users that fit their needs better than the choices users would have made on their own. This is due to, among other things, the vast amount of data that AI systems can analyze and the correlations they can reach that are beyond the capabilities of the human mind.¹⁵⁵ Similarly, AI systems do not suffer from human limitations, such as a reduced capacity for decision-making when handling an abundance of choice.¹⁵⁶ Can the predictive model of Lex AI thus enable AI to also determine for its users what their best interests *should* be? In other words, can Lex AI “derive an ‘ought’ from an ‘is’”?¹⁵⁷

There are many reasons why the answer is “no.” These reasons pertain both to Lex AI's decision-making process and the manner in which it is used. Two sets of reasons stand out. One, more general set of reasons, arises from the shift to automated decision-making processes and is grounded on the

151. See, e.g., Ryan Calo, *Digital Market Manipulation*, 82 GEO. WASH. L. REV. 995 (2014).

152. See generally DAVID S. EVANS & RICHARD SCHMALENSSEE, *MATCHMAKERS: THE NEW ECONOMICS OF MULTISIDED PLATFORMS* (2016) (exploring the business strategies of “matchmakers,” platforms who connect members for certain purposes, but whose business model focuses not on the services or products sold to the group through the platform but rather on the pool of members itself); David S. Evans, *Attention Platforms, the Value of Content, and Public Policy*, 54 REV. INDUS. ORG. 775, 777–78 (2019) (describing the basic economics of attention platforms).

153. See GILLESPIE, *supra* note 90, at 40–44 (describing how platforms curate content tailored to draw users, who pay with their attention to advertising); see generally TIM WU, *THE ATTENTION MERCHANTS: THE EPIC SCRAMBLE TO GET INSIDE OUR HEADS* (2016) (describing the rise of a new industry, one of users' attention, and explaining how it is used to drive purchase decisions).

154. For instance, system operators could tweak newsfeed or content recommendation algorithms so that radical and potentially damaging—albeit not illegal—content is pushed down the list of rankings. This would render it more likely that most users would never even see such content.

155. Sunstein, for instance, argues that such a paternalistic approach might be justified when applied in order to correct a behavioral bias or a systematic mistake, in which case it may actually promote people's autonomy. See Sunstein, *supra* note 120, at 437–38.

156. See Gal, *supra* note 104, at 72; see also Chagal-Feferkorn, *supra* note 107, at 144.

157. Joshua P. Davis, *Legality, Morality, Duality*, 1 UTAH L. REV. 55 (2014).

assumption that it is possible to define a person's "best interest" in the absence of an active choice. Defining a person's "best interest" can be based on various parameters and values. Even if Lex AI can accurately signal people's own preference, this may not necessarily indicate their best interest. For instance, individuals may act against their best interests due to favoring short-term preferences (e.g., by eating junk food) over long-term benefits (e.g., eating healthy).¹⁵⁸ People may also misinterpret their choices' true value¹⁵⁹ or the measures they need to take to maximize their preferences.¹⁶⁰ Moreover, decisions regarding "best interests" may entail open-ended questions involving contradicting interests and values and requiring judgment that may not be reduced to numerical ranking. In such cases, Lex AI may simply not be up for the task.¹⁶¹

Another set of reasons why Lex AI might fail to maximize users' best interest arises from core features that characterize machine learning systems: one reason is associated with the way ML involves learning from past instances (i.e., *path-dependency*) and other reasons have to do with the process by which the system generates its decisions (i.e., *optimization function* and *exploration-exploitation trade-offs*).

a) Path-dependency

As discussed in Section III.B, Lex AI reaches its decisions based on, among other things, information on users' past behavior. The correlation between past actions and best interests, however, is dubious. As noted, since individuals often act against their own best interests, relying on data that reflects past behavior may lead to undesired outcomes.¹⁶² This is exacerbated by AI

158. Anne van Aaken, *Judge the Nudge: In Search of the Legal Limits of Paternalistic Nudging in the EU*, in *NUDGING AND THE LAW, A EUROPEAN PERSPECTIVE* 89–91 (Alberto Alemanno & Anne Lise Sibony eds., 2015). For a discussion on cases where people's preferences are not in line with the public interest and when the state ought to attempt to change such preferences see Porat, *supra* note 7.

159. Oren Bar-Gill, *Algorithmic Price Discrimination: When Demand Is a Function of Both Preferences and (Mis)Perceptions*, 86 U. CHI. L. REV. 217, 244–45 (2018).

160. van Aaken, *supra* note 158, at 89–91.

161. For example, what constitutes an individual's best interest may vary with different moral frameworks when choosing between alternatives. One choice may be predicted to produce a certain gain for the individual; a different choice may produce a greater gain to the individual but involve a white lie; or an utterly different choice may produce the greatest gain but to a third person. AI systems, which lack humans' moral sense, may try to erroneously synthesize different frameworks and reach meaningless results. See Joshua P. Davis, *AI, Ethics, and Law: A Possible Way Forward*, in *ARTIFICIAL INTELLIGENCE AND PRIVATE LAW: GLOBAL PERSPECTIVES* (forthcoming, 2022).

162. See van Aaken, *supra* note 158, at 89–91.

decision-making's recursive nature, which factors in previous predictions made by the system through its feedback-loop mechanism. Thus, the system might not only perpetuate less favorable choices, but it could also be shape by and create them. For instance, if the system once mistakenly predicted that the user was interested in racist content and, despite a lack of interest and due to the nudging effect, the user clicked through to the racist content, the system might forever offer the user similar content, potentially intensifying the consumption of more radical content.¹⁶³

b) Optimization function

AI systems are tasked with reaching the best available predictions, which are determined based on the optimization of a predefined objective function.¹⁶⁴ The optimization function of a driverless vehicle, for example, may strive to maximize objectives such as safety, driving comfort, and speed—with the latter constrained by law. At the same time, it may endeavor to minimize fuel consumption and other environmental externalities.¹⁶⁵ Often, the different values may not reconcile or may even be contradictory to each other.¹⁶⁶ The system's designer determines which values to try to minimize or maximize and how much weight to attach to each of them.

Importantly, the decision of how to shape the optimization function and which values would affect it is reserved for the operators, or designers, of the system. The system's users are generally not part of the process for determining optimization criteria. Therefore, the users cannot affect a central element of how Lex AI norms are shaped. This may carry important implications for Lex AI's legitimacy. Although users' data provide an input into the decision-making process, the outcome of the process would be determined

163. This is in fact claimed to have been the case with respect to YouTube, whose algorithms allegedly presented alt-right content to non-alt-right viewers and, due to the system's recursive nature, have continued to "feed" users such content. Users who were shown this content took active part in alt-right movements and were reportedly "brainwashed" by YouTube. See Kevin Roose, *The Making of a YouTube Radical*, N.Y. TIMES (June 8, 2019), <https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html>.

164. See David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS L. REV. 653 (2017); OPTIMIZATION TECHNIQUES FOR PROBLEM SOLVING IN UNCERTAINTY (Surafel Lulseged Tilahun & Jean Medard T. Ngotchouye eds., 2018).

165. See YUMING GE, XIAOMAN LIU, LIBO TANG & DARRELL M. WEST, CTR. FOR TECH. INNOVATION AT BROOKINGS, SMART TRANSPORTATION IN CHINA AND THE UNITED STATES (2017).

166. Increasing speed, for example, might come at the expense of consuming more energy and causing more damage to the environment. Additional safety measures may come at the expense of driver comfort.

based on the optimization criteria the system and its operators adopt. Lex AI's outcome could hardly be said to reflect users' choice simply because their data was used as an input.¹⁶⁷

c) Exploration—Exploitation Tradeoffs

Another reason that the optimization function, which mathematically defines the system's goal, may not be fully aligned with each user's best interests is that systems often factor in operational requirements. Such considerations of the common good may sometimes result in sacrificing certain users. Consider, for example, a traffic jam which requires Waze to redirect drivers to alternative roads. Hypothetically, if the app directs all drivers to the shortest route, that route may become congested, and the app would no longer serve the best interest of any of its users. Therefore, when all drivers are using Waze, Waze has to reach a social optimum for all drivers, even if it runs contrary to the optimum for any single driver.

AI systems' operational constraints may present another prioritization of the common good at the expense of the best interests of individual users. AI systems maintain a known tradeoff. On the one hand, the system is programmed to exploit information already acquired and presently available in order to reach the optimal choice. On the other hand, the system needs to gather new information and learn about new alternatives that could prove favorable to the best presently available options.¹⁶⁸ In other words, in order to improve its ability to offer better choices, the system must explore new options, despite the fact that these new options may lead to suboptimal outcomes for certain users. To better serve its users, a restaurant recommendation system, for example, will need to obtain reviews on establishments that were not recently visited. Thus, although the system may

167. Moreover, and as discussed earlier, although centralized AI systems account for the users' best interests or preferences, the systems might actually be used to optimize additional objectives. Matching algorithms—such as those matching consumers with delivery services or vacation rentals—may consider users' desires and needs, but the system would still strive to maximize their operators' profitability. To reach optimal allocation of rentals, for example, the system may recommend less attractive rentals to the users whose profiles indicate they would agree to such options, while leaving the better rentals only for users the system identifies as picky.

168. See Gal Bahar, Omer Ben-Porat, Kevin Leyton-Brown & Moshe Tennenholtz, *Fiduciary Bandits*, 37 PROCS. INT'L CONF. ON MACHINE LEARNING 518, 518 (2020), <http://proceedings.mlr.press/v119/bahar20a/bahar20a.pdf> [hereinafter *Fiduciary Bandits*]; Gal Bahar, Omer Ben-Porat, Kevin Leyton-Brown & Moshe Tennenholtz, *Learning under Invariable Bayesian Safety* (Arxiv, Working Paper No. 2006.04497, 2020), <https://arxiv.org/abs/2006.04497>.

generally aim to recommend restaurants to its users that are known to match their profile, occasionally, in order to learn and improve, the system may choose to steer users in the direction of choices whose utility value for the individual user is unknown.¹⁶⁹ By directing users to explore alternatives whose utility is unknown, the system is “intentionally providing . . . sub-optimal recommendations” to some of its users.¹⁷⁰

3. *The Decline of the Deliberative Space*

Returning to Mill’s critique of the “good despot,” Mill did not merely focus on the feasibility of accumulating and processing sufficient information to enable efficient governance. Mill also argued that such a governance structure lacks the sort of participatory component essential—not only for the sake of efficiency—but also for moral justifiability.¹⁷¹ Consequently, even if Lex AI were capable of generating full and accurate data on public needs and preferences and could accurately calculate the overall public good, Mill’s critique suggests that lacking a participatory dimension would render it morally inferior.

The democratic ideal of self-governance assumes access to information and the right to free and informed deliberation so that citizens may individually express their autonomy and collectively decide their common destiny.¹⁷² Participation in crafting social norms takes different shapes and forms, ranging from directly voting by referendum on policy initiatives to electing representatives that would promote a particular governmental agenda to participating in public discourse and influencing the formation of norms. According to a “participatory theory of governance,” wide participation in public discourse is a vehicle for enabling self-governance and constructing

169. The same is true for Waze. While the system’s objective is to minimize user’s travel time, it might nevertheless recommend routes which are not necessarily the fastest, to obtain real-time information on them. In other words and with respect to routes that Waze lacks up to date information on (and thus does not know if are the fastest or not), Waze may need to direct drivers to these routes despite potentially causing drivers longer travel time than necessary.

170. *Fiduciary Bandits*, *supra* note 168, at 1.

171. MILL, *supra* note 27, at 68 (“Still more salutary is the moral part of the instruction afforded by the participation of the private citizen . . . From these accumulated considerations it is evident that the only government which can fully satisfy all the exigencies of the social state is one in which the whole people participate; that any participation, even in the smallest public function, is useful; that the participation should everywhere be as great as the general degree of improvement of the community will allow; and that nothing less can be ultimately desirable than the admission of all to a share in the sovereign power of the state”).

172. Ellen P. Goodman, *Digital Information Fidelity and Friction*, KNIGHT FIRST AMENDMENT INST. (2020), <https://knightcolumbia.org/content/digital-fidelity-and-friction>.

democratic legitimacy.¹⁷³ In other words, in democracies, the aggregated will of the people is reflected in government by their participation—specifically, by electing the ruler and, more generally, by participating in the discursive formation of collective norms to which individuals will be subject. The governance of Lex AI does not reflect active participation. Although it could be argued that the collection of user preferences and the usage of their aggregation to shape personalized norms is a form of passive participation, such participation does not reflect individuals' choices and potentially not even their true preferences, as discussed in Section III.B. Moreover, the outcomes of Lex AI cannot signal the aggregated will of users, as explained in Section III.C.2.

Furthermore, the mechanisms available for public participation in a democratic setting involve an explicit articulation of norms and underlying values that enable public discussions and oversight. Participation in political debates, accordingly to Mill's political theory, could strengthen citizens' bond to the community and may expand their conception of the common good beyond the narrow lens of their private affairs.¹⁷⁴

The deliberation over these norms and values is also possible in other venues. The law, for example, resolves tensions between contradictory interests through various layers of conflict resolution that enable deliberation and negotiation of norms through multiple social institutions.¹⁷⁵

Naturally, Lex AI also involves trade-offs between conflicting interests and values. But it fails to provide similar space for deliberating upon those trade-offs. Unlike governance by social or legal norms—which shape behavior by conveying information on right and wrong, often followed by a sanction or reward¹⁷⁶—Lex AI shapes behavior without resorting to the communicative

173. See generally Robert Post, *Reconciling Theory and Doctrine in First Amendment Jurisprudence*, 88 CALIF. L. REV. 2355 (2000).

174. Citizen participation, Mill argues, has an educational value that could transform the participating citizens. A participating citizen “is called upon, while so engaged, to weigh interests not his own; to be guided, in case of conflicting claims, by another rule than his private partialities; to apply, at every turn, principles and maxims which have for their reason of existence the general good He is made to feel himself one of the public, and whatever is for their benefit to be for his benefit.” MILL, *supra* note 27, at 68; see also Alex Zakaras, *John Stuart Mill, Individuality, and Participatory Democracy*, in J.S. MILL'S POLITICAL THOUGHT: A BICENTENNIAL REASSESSMENT 200, 207–10 (Nadia Urbinati & Alex Zakaras eds., 2007).

175. See Elkin-Koren & Perel, *supra* note 84.

176. See Katz, *supra* note 48, at 1749 (“Norms and rules, whether publicly or privately created, embody and convey information. They cannot be followed unless information is transmitted regarding their substantive content; they cannot be enforced unless information is

nature of norms. Norms fashioned by AI systems would be shaped over time, depending on the system design as well as the type of data fed into the system through its continual use. Such processes do not explicitly involve any value-based choices, and they are often not transparent but hidden behind a veil of technical details. Consequently, they do not facilitate the development of new conceptions of values and trade-offs in an intelligible manner.¹⁷⁷

Furthermore, changes to the norms might be implicit in the system's learning and may not necessarily reflect any desirable social optimum or social choice. This way in which changes to norms occur does not facilitate deliberation over values and the trade-offs involved in such decision-making processes. Consider, for instance, the governance of speech. Legal norms shape speech through explicit rules and principles by offering a definition of what speech is and the limitations to which it is subject.¹⁷⁸ In Lex AI, by contrast, predictions on whether a particular speech act should be deemed illicit are not the result of conscious deliberation on underlying free speech principles but instead may depend on many dynamic variables: whether the content has triggered a computational threshold; whether similar content has triggered the system before; whether third parties have flagged the content or similar content; who flagged the content; and how often any of these things have occurred.¹⁷⁹ Shifting from governing speech by law to governing speech by Lex AI lacks an essential social decision-making mechanism for developing, contesting, and socially negotiating speech norms.¹⁸⁰ Lex AI also fails to reveal not only the trade-offs but also the procedures by which such tradeoffs between conflicting interests are decided.

All in all, the absence of discursive dimension in Lex AI decision-making processes weakens its legitimacy as a form of collective self-governance.

transmitted regarding who has obeyed them, who has violated them, and who is to impose any associated punishment or reward.”).

177. In their book, Frischmann and Selinger refer to the following quote by Baruch Spinoza: “Experience teaches us no less clearly than reason, that men believe themselves free, simply because they are conscious of their actions, and unconscious of the causes whereby those actions are determined.” BRETT FRISCHMANN & EVAN SELINGER, RE-ENGINEERING HUMANITY 222 (2018). According to Frischmann and Selinger, “[t]hrough not the conventional interpretation, Spinoza also might be read to suggest that what we are conscious of shapes what we believe and blinds us to the hidden complexities of multiple interdependent causes and contingencies that dramatically shape—and in that more limited sense determine—our beliefs, desires, and actions.” *Id.*

178. See Elkin-Koren & Perel, *supra* note 84.

179. See *id.*

180. See *id.*

IV. LEX AI AS A *SUI GENERIS* ORDERING SYSTEM: LEGAL IMPLICATIONS

The distinction between public and private ordering is significant in several policymaking contexts. For example, the U.S. Constitution restrains the use of governmental power (i.e., state power), but it generally does not apply to private ordering by private actors.¹⁸¹ Policy interventions such as ensuring that norms reflect the parties' autonomous choice are prevalent and necessary when private and not public ordering is concerned.

The Article, however, has demonstrated that Lex AI introduces a new type of governance, one which neither fits neatly under the public ordering nor private ordering classifications. As demonstrated by our analysis, Lex AI can hardly be viewed as a bottom-up process whereby norms are crafted and undertaken by the stakeholders to which they apply. Its predictions are based on a centralized decision-making process that aggregates personalized data to predict outcomes for individuals based on its optimization function. Bypassing autonomous choice by users, Lex AI may look more like a distinctive type of collective action mediated by algorithms rather than self-governance. As a result, although overcoming informational barriers by offering efficient means to manage, organize, and analyze data, Lex AI introduces serious challenges to personal choice and does not entertain the legitimacy of private ordering.

At the same time, however, Lex AI cannot be viewed as strictly top-down governance. Lex AI is not tightly controlled by those who designed it or deployed it. It shapes behavior by data-driven statistical predictions. Consequently, multiple sources of data shape its output (i.e., predictions) and adaptive learning function. Therefore, they are more distributed, more dynamic, and less predictable than command-and-control governance.

All in all, Lex AI introduces a *sui generis* governance which is neither purely centralized nor distributed. This observation may carry several implications for policymakers seeking to define the scope of regulatory intervention in Lex AI and the appropriate measures of intervention.

First, focusing on Lex AI's private ordering affordances assures that user choices accurately reflect their true preference. This Article has shown that Lex AI does not provide a reliable signaling of people's preferences and choices. The way preferences are inferred and the recursive process by which Lex AI shapes norms and behavior may result in predictions that fail to reflect

181. For instance, freedom of expression as a constitutional right ensures that many governmental attempts to restrict speech would be subject to strict scrutiny under the First Amendment.

individuals' true preferences and may generate inefficiencies. Since private ordering presumes that the parties know best for themselves, public intervention seeks to verify that parties' choices are informed by setting disclosure obligations or prohibiting misleading practices.¹⁸² Under Lex AI, where parties' preferences are recursively inferred, other measures might be necessary to assure parties' ability to make choices that reflect their true preferences. For example, a periodic report may be sent to Lex AI users, detailing their various classifications in the system (e.g., fans of horror movies, having an *X* percent aversion to certain types of violence, people who care much about the watching habits of their peers, etc.) and allowing them to opt out of classifications they believe are incorrect or otherwise undesirable. Another option would be to have the system flag for the user past choices or behaviors that had a larger than average impact on the system's recursive decision-making process with respect to the specific user, and ask the user to provide input on how the system should weigh this information.

Second, public policy measures related to public ordering considerations may be adjusted to account for the fact that the decisions are being made by Lex AI's rule. Unlike governance by legal norms which conveys information on right and wrong (often followed by a sanction or reward),¹⁸³ Lex AI shapes behavior without resorting to the communicative nature of norms. Unlike governance through legal norms or rule-based code, Lex AI cannot be analyzed once, *ex ante*, for extracting the values, choices, and trade-offs it embeds. Its adaptive function is driven by data, which renders this type of governance inherently dynamic. Moreover, as norms embedded in the system are opaque and nondiscursive, this governance form fails to facilitate explicit deliberation on social norms. The decline of a discursive dimension, which is essential for determining the societal values that inform public policy, may require policymakers to consider new types of interventions to enable more public deliberation over the trade-offs embedded in Lex AI design. For example, when AI systems driven by common good considerations perform the explore and exploit trade-off, regulation can assure an equal distribution of the "explore" alternative among users.

Third, Lex AI decisions may require public policy to treat Lex AI as an ecosystem. The behaviors and interactions between various types of (sometimes unidentified) entities shape Lex AI decisions. Each affects the adaptive learning of the centralized system and, consequently, the decisions the system generate. Treating Lex AI as a data ecosystem rather than a rule-

182. *See, e.g.*, Shmuel I. Becher & Oren Bar-Gill, *Consumer Protection*, in *THE ECONOMIC APPROACH TO LAW* (Uriel Procaccia ed., 2012).

183. *See* Katz, *supra* note 48, at 1749.

based design would assist in focusing attention on the interaction between the different actors when considering legal tools to mitigate potential harms the system generates. For instance, if policymakers seek to limit certain harms Lex AI causes, they should bear in mind that code alone does not necessarily determine the outcomes they wish to mitigate. Policymakers should consider that such outcomes result from data originated from various sources. This could be an accumulated effect of crowds of users or data tweaked by strategic players such as governments, adversary nongovernmental entities, or competing corporate players.

Consider the vulnerability of algorithmic content moderation systems on social media to misuse by strategic players. Governments, for instance, may seek to slow down the spread of certain content on social media platforms¹⁸⁴ by filing removal requests using social media notice and take down procedures.¹⁸⁵ Due to the recursive feedback loop, content subsequently flagged as illicit is likely to be fed back into the model so that it will be detected the next time the system runs, and it thus may impact not simply current removal decisions but also future removals of similar content. Consequently, the systematic issuing of removal requests may affect the technical definitions of what is considered illicit content and may tilt the AI based filters towards governmental perception of what counts as illegal.¹⁸⁶

184. In the United Kingdom, for instance, the Counter-Terrorism Internet Referral Unit (CTIRU) identifies content that breaches the terms of service of social media platforms and requests that they remove the content on a voluntary basis. Established in 2010 by the Association of Chief Police Officers (ACPO) (a nongovernmental body later replaced by the National Police Chiefs' Council (NPCC) (also a private company which coordinates antiterrorist efforts)) and run by the police, this initiative acts to remove terrorist material from digital platforms. CTIRU focuses on UK-based materials but also compiles lists of URLs for material hosted outside the United Kingdom for the relevant service providers to block. *Counter-Terrorism Internet Referral Unit*, OPEN RIGHTS GRP., https://wiki.openrightsgroup.org/wiki/Counter-Terrorism_Internet_Referral_Unit (last visited Feb. 2, 2022). Similarly, in Israel the Cyber Unit at the State Attorney's Office deployed a comprehensive and elaborate enforcement of speech regulation through digital platforms. The unit systematically files removal requests with digital platforms, targeting allegedly illegal content, such as postings that instigate violence against judges or other public servants, threats against minors, or materials inciting terrorism. *About the Cyber Unit*, GOV.IL (Nov. 5, 2021), <https://www.gov.il/en/departments/general/cyber-about>.

185. Requests are made based on the platforms' own content moderation policy, as reflected in its Terms of Use (ToU), community standards or community guidelines. Platforms generally exercise discretion over whether such content violates their ToU and decide which actions to take regarding such content.

186. This vulnerability of the platform's content moderation systems to strategic flagging by governments, is demonstrated by Chinese government use of YouTube's rules to silence human rights activists. YouTube recently removed an activist channel that collected and

This perspective may highlight the potential of particular data to cause harm in ways which were unintended and unpredictable by the original designer or by those who deploy the system. Critically, however, treating Lex AI as an ecosystem where design and deployment decisions pertaining to the system might be affected by different stakeholders should not be understood as a vehicle for shielding different stakeholders from liability for their design choices—including those concerning which data to collect and models to deploy. A better understanding of the interactions between stakeholders may assist in tailoring better regulatory and legal measures that will respond to the big picture and not target only a small subset of potential concern.

V. CONCLUSION

A further inquiry into the way Lex AI governs may carry important implications for assessing its potential efficiency and determining the scope of its legitimacy. Lacking some key features of private ordering—like informed, voluntary, and self-imposed choice on the one hand, while on the other hand, also lacking explicit and coherent procedures for deciding societal best interest—Lex AI could be thought of as a *sui generis* type of governance. This, we argue, requires fresh thinking on the role of law and on the scope and type of possible legal intervention in relation to Lex AI.

published video testimonies from family members of people imprisoned in China's internment camps in Xinjiang. The channel was removed following mass flagging campaigns the Chinese and Kazakh governments allegedly orchestrated. Supporters were instructed to follow a video explaining how to flag the videos en masse in order to force YouTube to take them down. Eileen Guo, *How YouTube's Rules are Used to Silence Human Rights Activists*, MIT TECH. REV. (June 24, 2021), <https://www.technologyreview.com/2021/06/24/1027048/youtube-xinjiang-censorship-human-rights-atajurt/>.

TECHNOLOGY LAW AS A VEHICLE FOR TECHNOLOGY JUSTICE: STOP ISP THROTTLING TO PROMOTE DIGITAL EQUITY

By Catherine J.K. Sandoval[†]

ABSTRACT

Society has shifted to bandwidth intensive internet use to protect public health, sustain economic participation and educational attainment, and support critical infrastructure services. This shift made robust, open, broad-based internet network (ROBIN) access essential to public safety and societal resilience. This Article examines Internet Service Provider (ISP) throttling policies, an underexplored area of the digital divide, net neutrality, and consumer protection legal and policy debate. Many ISPs use contract terms and software to slow users to 2G speeds, more commonly used in the early 1990s, after internet use commensurate with a few days or a week or two of digital work or school. The difference between open internet access and throttled access is the ability to access school, work, or a telemedicine appointment through videoconferencing platforms and other applications.

This Article theorizes infrastructure law and policy as a legal superstructure that either reifies vulnerability and inequity or, if constructed differently, supports equity and technology justice. To build the legal architecture necessary to foster ROBIN access critical to education, health, economic, and civic opportunities, this Article calls for Federal Communications Commission (FCC) and Federal Trade Commission (FTC) examination of ISP throttling policies and disclosures. It calls on ISPs, as part of their commitments to diversity, inclusion, and equity, to drop the practice of throttling users to speeds that make contemporary internet uses unavailable. It argues that stemming ISP throttling is critical to equity, opportunity, public health, and public safety.

DOI: <https://doi.org/10.15779/Z38ST7DX70>

© 2021 Catherine J.K. Sandoval.

[†] Catherine J.K. Sandoval, Associate Professor, Santa Clara University School of Law (SCU Law); Former Commissioner, California Public Utilities Commission; and Director, Broadband and Infrastructure Institute @ Santa Clara University School of Law (BBIC). Thanks to the participants in and organizers of the U.C. Berkeley Symposium, Technology Law As a Vehicle for Anti-Racism. Special thanks to Tejas Narechania, Olivier Sylvain, Ari Q. Fitzgerald, and Ernesto Falcon for their comments on an earlier version of this paper and their contributions to the symposium. Thanks to BBIC founder Allen S. Hammond, IV, and to SCU Law students Kasey Kagawa, Robert Murillo, and Rosa Rico for their collaborative work on the BBIC's comments to the FCC examining throttling as a net neutrality and safety issue on remand from *Mozilla v. FCC*. Special thanks to my husband, Steve Smith, for his support, encouragement, and commitment to equity.

TABLE OF CONTENTS

I.	TECHNOLOGY JUSTICE: INTRODUCTION	964
A.	INTERNET SERVICE PROVIDER THROTTLING UNDERMINES PUBLIC HEALTH, PUBLIC SAFETY, AND THE PUBLIC INTEREST	964
II.	THE INTERNET MEDIATES ACCESS TO RESOURCES AND THE FUTURE	968
III.	LEGAL FRAMEWORK FOR FCC REGULATION TO PROMOTE SERVICE TO ALL AMERICANS IN THE PUBLIC INTEREST	979
A.	PROMOTING WIRELESS AND WIRELINE SERVICE TO ALL AMERICANS WITHOUT DISCRIMINATION	979
B.	PROMOTING ADVANCED TELECOMMUNICATIONS SERVICES AND THE OPEN INTERNET’S VIRTUOUS CIRCLE OF INNOVATION	982
C.	CENTERING THE PUBLIC IN PUBLIC SAFETY	992
D.	THE EMERGENCY BROADBAND BENEFIT PROGRAM	997
IV.	FTC ACT DECEPTIVE CONDUCT PROSCRIPTIONS.....	999
V.	THROTTLING INTERNET ACCESS LIMITS EQUITY AND RISKS PUBLIC HEALTH AND SAFETY.....	1002
A.	INTERNET ACCESS NEEDS TO ZOOM PAST THE FCC’S ADVANCED SERVICES DEFINITION.....	1002
B.	THROTTLING TO THE ’90S	1009
VI.	RECOMMENDATIONS AND CONCLUSION	1014

I. TECHNOLOGY JUSTICE: INTRODUCTION

A. INTERNET SERVICE PROVIDER THROTTLING UNDERMINES PUBLIC HEALTH, PUBLIC SAFETY, AND THE PUBLIC INTEREST

The coronavirus pandemic commencing in March 2020¹ marked a flash cut to digital education, work, and health services. As schools, colleges, employers, healthcare providers and other institutions switched to online services to protect public health and safety during the COVID-19 pandemic, new digital divide fissures opened. Videoconferencing applications such as Zoom and WebEX became necessary for telemedicine, class discussions, and many jobs. At the same time, Internet Service Provider (ISP) throttling

1. See *WHO Director-General’s Opening Remarks at the Media Briefing on COVID-19 - 11 March 2020*, WORLD HEALTH ORG. (Mar. 11, 2020), <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>.

policies (slowing user speeds after certain data consumption or network condition triggers are met)² made those applications inaccessible to throttled consumers. ISP throttling disables user access to video-based telemedicine, tele-education, and many facets of online work for hours, days, or weeks at a time.³ Throttling is not merely a technical, regulatory, or legal issue; it creates public safety, health, resiliency, and equity risks. To build the legal architecture necessary to foster robust, open, broad-based internet network (ROBIN) access critical to education, health, economic, and civic opportunities, this Article calls for Federal Communications Commission (FCC) and Federal Trade Commission (FTC) examination of ISP throttling policies and disclosures.

For more than a decade, some ISPs have throttled users' internet service on the basis of various contractual triggers or conditions. Throttling may occur based on what the ISP deems "excessive usage" by a consumer, defined as higher-than-average usage for plan subscribers, even if that usage is below the plan's data cap or the consumer pays for an "unlimited" data plan.⁴

In 2021, several ISPs offering prepaid wireless services and some wireline ISPs triggered contract-based throttling after subscriber internet use commensurate with a few days or a week or so of digital work or school.⁵

2. Tyler Cooper, *How to Tell if Your Internet is Being Throttled*, BROADBANDNOW (Aug. 10, 2021), <https://broadbandnow.com/guides/am-i-being-throttled/>; see also Fed. Trade Comm'n v. AT&T Mobility LLC, 883 F.3d 848, 850 (9th Cir. 2018) (affirming Federal Trade Commission (FTC) jurisdiction to bring a deceptive conduct complaint against "AT&T Mobility's 'data throttling'—a practice by which the company reduced customers' broadband data speed without regard to actual network congestion").

3. See *infra* Section V.B.

4. Catherine J. K. Sandoval, *Disclosure, Deception, And Deep-Packet Inspection: The Role of The Federal Trade Commission Act's Deceptive Conduct Prohibitions in the Net Neutrality Debate*, 78 FORDHAM L. REV. 641, 688 (2009) [hereinafter Sandoval, *Disclosure, Deception, and Deep-Packet Inspection*] (reporting in 2009 that cable-based ISP Time Warner prohibits 'use of excessive bandwidth' in its operator acceptable use policy without defining what constitutes excessive use"); *id.* at 698–99 ("As of August 2009 FIOS did not impose bandwidth limits, but its acceptable use policy prohibits subscribers from generating excessive internet traffic, a level it does not define."); Jon Brodtkin, *Cox Slows Internet Speeds in Entire Neighborhoods to Punish any Heavy Users*, ARS TECHNICA (June 8, 2020), <https://arstechnica.com/tech-policy/2020/06/cox-slows-internet-speeds-in-entire-neighborhoods-to-punish-any-heavy-users> ("Cox Communications is lowering internet upload speeds in entire neighborhoods to stop what it considers 'excessive usage,' in a decision that punishes both heavy internet users and their neighbors.").

5. See Comments of the Broadband Institute of California @ Santa Clara University School of Law (BBIC), In the Matter of Restoring Internet Freedom (WC Docket Nos. 17-108, 17-287, 11-42), at 7–8 (Apr. 20, 2020), <https://ecfsapi.fcc.gov/file/104211478729214/BBIC%20Comments%20FCC%20Net%20Neutrality%20Mozilla%20remand%20final.pdf> [hereinafter BBIC, *Comments on Mozilla Remand*].

Through contract, several ISPs reserved the right to slow users to 2G speeds, which were more commonly used in the early 1990s.⁶ Most ISPs no longer operate 2G networks. Instead, ISPs use software to slow users who hit throttling triggers and maintain users at throttled speeds.⁷ Software-enforced throttling may last for minutes, hours, days, or weeks at a time, regardless of network congestion.

ISPs may throttle a household, or even a neighborhood,⁸ effectively disabling tele-education, telemedicine, or remote work. ISP throttling often restricts the user's speed below the threshold required for contemporary internet uses. The difference between open internet access and throttled access is the ability to access school, work, or a telemedicine appointment through videoconferencing platforms and other applications. Millions of American individuals, families, and neighborhoods subject to ISP throttling practices bear a heavy burden from loss of meaningful internet access. Society suffers from ISP throttling as education, health, civic participation, and employment options are narrowed for a large, but unreported, number of Americans. Institutions responsible for delivering health, educational, critical infrastructure, and civic engagement services face ISP-imposed reliability and internet access constraints that narrow service delivery, public health, and safety options.

During the COVID-19 pandemic, society shifted to bandwidth-intensive internet use to sustain economic participation and educational attainment while protecting public health. Videoconferencing and video applications use more bandwidth (internet capacity) than applications such as email and are extremely sensitive to latency (delays in communicating internet signals). ISP-induced slowdowns may render those internet resources unavailable. Use of videoconferencing services and other bandwidth-intensive applications continued during the first two years of the COVID-19 pandemic as a backup plan for and complement to in-person services. As internet use continues to evolve, such bandwidth-intensive services will continue to attract innovative internet use.

The internet increasingly mediates access to resources and services. Ensuring that people have robust and open internet access is critical to public

6. *See infra* Part V.

7. *See* Stetson Doggett, *How Fast are Capped 2G Speeds? LTE vs 3G vs 2G Data Speed Test*, BEST PHONE PLANS (Mar. 14, 2021), <https://www.bestphoneplans.net/news/2g-speed-test>.

8. Rebecca Lee Armstrong, *How can I Tell if My Internet is Being Throttled by My ISP?*, HIGH SPEED INTERNET.COM (Feb. 4, 2020), <https://www.highspeedinternet.com/resources/how-can-i-tell-if-my-internet-is-being-throttled-by-my-isp> (“During times of heavy internet use in a single area, ISPs sometimes throttle everyone’s internet in that area.”)

safety, public health, education, and economic and civic opportunity. Law, policy, and theory should rectify rather than reify technological and concomitant social inequality.

This Article does not relitigate the net neutrality debate, aspects of which four of my previous articles analyzed.⁹ This Article focuses on the pressing problem of ISP throttling to 2G speeds, particularly for prepaid mobile customers, and inadequate disclosure of those practices.

Part II of this Article analyzes the internet's increasing integration into the economy and civic life and argues that the internet increasingly mediates access to resources, underscoring the imperative of sound legal regulation of this sector. Part III analyzes the legal framework for FCC regulation that requires the FCC to promote access to advanced telecommunications services for all Americans.

Part IV provides a brief overview of Federal Trade Commission Act (FTCA) deceptive conduct proscriptions. It examines ISP throttling in the context of the FTC's 2014 complaint against AT&T Mobility alleging that AT&T's practice of throttling internet access wireless plans that were advertised as "unlimited," without clear disclosures about when and how such throttling would be triggered or explanation of its consequences, violated FTCA proscriptions against deceptive conduct in interstate commerce.¹⁰ The FTC and AT&T settled that complaint in 2019 through a permanent restraint against AT&T making any express or implied representations "about the amount or speed of mobile data, including that the mobile data is unlimited, without disclosing, clearly and conspicuously and in close proximity to the representation, all material restrictions imposed by defendant."¹¹ ISP throttling, including AT&T's practices, raise compliance issues with the FTCA and the FTC-AT&T Mobility settlement.

Part V examines ISP internet access throttling executed through contracts whose terms provide insufficient disclosure to support consumer choice or contemporary communications needs. It argues that those disclosures fail

9. Sandoval, *Disclosure, Deception, and Deep-Packet Inspection*, *supra* note 4; Catherine J.K. Sandoval, *Net Neutrality Powers Energy and Forestalls Climate Change*, 9 SAN DIEGO J. CLIMATE & ENERGY L. 1, 81 (2018); Catherine J.K. Sandoval, *Net Neutrality Repeal Rips Holes in the Public Safety Net*, 80 U. PITT. L. REV. 953 (2019); Catherine J.K. Sandoval, *Cybersecurity Paradigm Shift: The Risks of Net Neutrality Repeal to Energy Reliability, Public Safety, and Climate Change Solutions*, 10 SAN DIEGO J. CLIMATE & ENERGY L. 91, 171 (2019).

10. Complaint for Permanent Injunction and Other Equitable Relief, *FTC v. AT&T Mobility LLC*, No. 3:14-cv-04785 (N.D. Cal. filed Oct. 28, 2014).

11. See Stipulated Order for Permanent Injunction and Monetary Judgment at 4, I(a), *FTC v. AT&T Mobility LLC*, No. 3:14-cv-04785 (N.D. Cal. Dec. 3, 2019) [hereinafter, *FTC AT&T Mobility Stipulated Order*].

FCC and FTCA standards, while ISP throttling interferes with critical communications needs and vital services.

Part VI concludes with recommendations for FCC and FTC examination of ISP throttling practices. This Article urges the FCC to examine ISP throttling under the FCC's transparency rule, 47 CFR 8.11, still standing after the FCC's 2018 net neutrality repeal, affirmed by the FCC in 2020.¹² It recommends FTC analysis of ISP representations about speeds and throttling under the FTC Act's deceptive conduct provisions and monitoring for compliance with the *AT&T Mobility-FTC* settlement.¹³

This Article urges ISPs to drop their throttling policies, consistent with ISP no throttling commitments¹⁴ and ISP diversity and inclusion policies. It encourages the FCC, states, academics, and the public to gather more information about ISP throttling practices. It argues for reframing FCC internet regulation to put the public at the center of the regulatory paradigm and recognize that communication by and between the polity is critical to public safety and the public interest. This framework is critical to development of ROBIN network access, investments in network expansion, support for internet users, the economy, public health, public safety, and our common future.

II. THE INTERNET MEDIATES ACCESS TO RESOURCES AND THE FUTURE

The internet is critical infrastructure that increasingly mediates access to educational, health, civic, and public safety resources, and the economy. “California Senate Bill 822 recognizes that [a]most every sector of California’s economy, democracy, and society is dependent on the open and

12. In the Matter of Restoring Internet Freedom, 35 FCC Rcd. 12328, 12329, ¶ 2 (2020) [hereinafter *2020 RIF Order*]; Restoring Internet Freedom, 83 Fed. Reg. 7852, 7852 (Feb. 22, 2018) [hereinafter FCC, *2018 RIF Order*]; Mozilla Corporation v. FCC, 940 F.3d 1, 49 (D.C. Cir. 2019) (upholding the FCC's adoption of a modified transparency rule under 47 U.S.C. § 257 which requires the FCC to consider “market entry barriers for entrepreneurs and other small businesses”); *id.* at 49 (upholding against APA challenge FCC transparency rule adopt to “keep entrepreneurs and other small businesses [as well as consumers] effectively informed of [broadband provider] practices so that they can develop, market, and maintain Internet offerings” (citing Restoring Internet Freedom Order, 33 FCC Rcd. 311, 439–42, ¶¶ 218–23 (2018))).

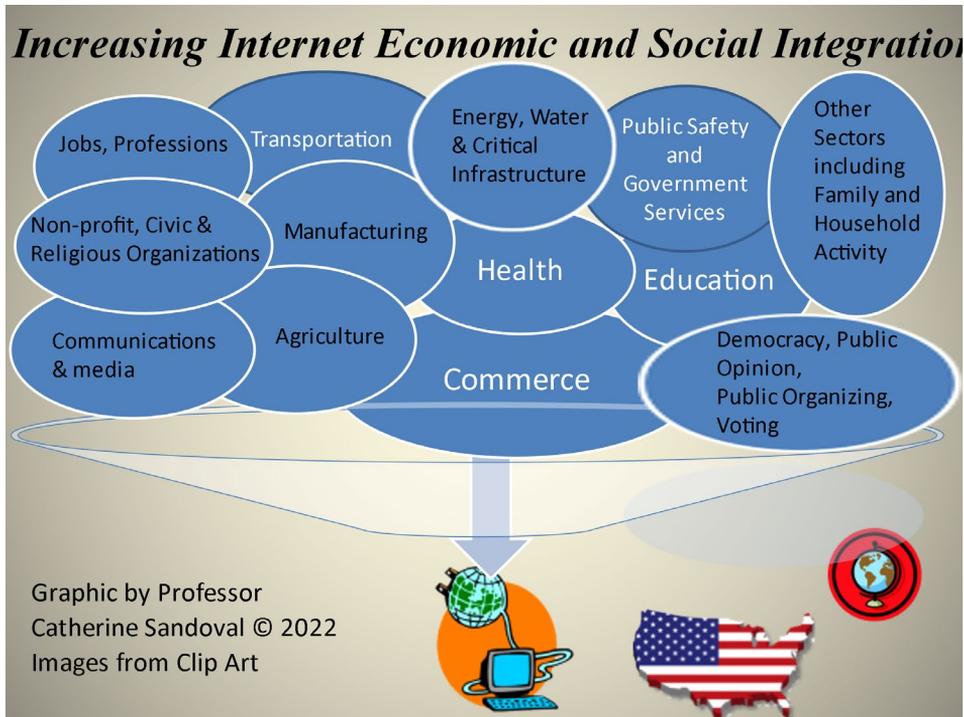
13. 15 U.S.C. §§ 41–77; see Sandoval, *Disclosure, Deception, and Deep-Packet Inspection*, *supra* note 4, at 666–94 (describing FTC deceptive conduct proscriptions, and defenses as applied to ISP conduct); *FTC AT&T Mobility Stipulated Order*, *supra* note 11, at 4.

14. *2020 RIF Order*, *supra* note 12, at 12349–50, ¶ 39.

neutral [i]nternet that supports vital functions regulated under the police power of the state,' including 'Utility services and infrastructure.'¹⁵

As illustrated by Figure 1 below, the internet is increasingly embedded into myriad economic and social sectors and activities. These include: jobs and professions; transportation; energy, water and critical infrastructure; public safety and government; non-profit, civic and religious organizations; manufacturing; commerce; agriculture; health; education; communications and media; democracy, public opinion, public organization and voting; and other sectors including family and household activity. Internet access increasingly filters economic, societal, and democratic resources, services, and engagement.¹⁶

Figure 1: The Internet's Integration into the Economy and Society



Lack of robust, open, and broad-based information and communications technology (ICT) frustrates access to services and resources, entrenches

15. Sandoval, *Net Neutrality Powers Energy and Forestalls Climate Change*, *supra* note 9, at 77.

16. MANUEL CASTELLS, *THE RISE OF THE NETWORK SOCIETY* xviii (2d ed. 2020); JAN VAN DIJK, *THE DIGITAL DIVIDE* 5 (2020); *Packingham v. North Carolina*, 137 S. Ct. 1730, 1735 (2017).

community poverty, and hobbles initiatives to connect data to action.¹⁷ “Digital exclusion forms a recursive process that undercuts safety and community resiliency.”¹⁸ Network exclusion burdens fall heavily on those excluded or under-included.¹⁹ Society bears the costs and responsibility for network exclusion.

Conversely, ICT investment and well-calibrated regulation enable opportunity and innovation for generations.²⁰ Infrastructure regulation creates the future’s physical and social architecture.²¹ Universal service theory highlights network expansion’s benefit to all users.²² “Universal service principles form the bedrock of communications policy, recognizing that the network is stronger as everyone is connected and served.”²³ The “universal service objective is founded on the concept that all subscribers to a telephone company’s basic service network benefit when another person joins that network. Therefore, the entire network is more valuable because of the addition of the new subscriber.”²⁴

17. See Catherine J.K. Sandoval, *Energy Access is Energy Justice, The Yurok Tribe’s Trailblazing Work to Close the Native American Reservation Electricity Gap*, in ENERGY JUSTICE, US AND INTERNATIONAL PERSPECTIVES 122 (Raya Salter, Carmen G. Gonzalez & Elizabeth Ann Kronk Warner eds., 2018); cf. WOODROW CLARK, SUSTAINABLE CITIES AND COMMUNITIES DESIGN HANDBOOK: GREEN ENGINEERING, ARCHITECTURE, AND TECHNOLOGY 121–22 (2d ed. 2018).

18. Catherine J.K. Sandoval & Patrick Lanthier, *Connect the Whole Community; Leadership Gaps Drive Disaster Vulnerability and the Digital Divide*, in TECHNOLOGY VS GOVERNMENT: THE IRRESISTIBLE FORCE MEETS THE IMMOVABLE OBJECT 1 (Lloyd Levine ed., 2022).

19. Ernest J. Wilson III, Sasha Costanza-Chock & Michelle Forelle, *A Provocation on Behalf of the Excluded*, in THE COMMUNICATION CRISIS IN AMERICA, AND HOW TO FIX IT 257–60 (Mark Lloyd & Lewis Friedland eds., 2016).

20. See *What Catastrophe Tells Us about Technology and Society*, in SHAPING TECHNOLOGY / BUILDING SOCIETY 7, 11 (Wiebe Bijker & John Law eds., 1994); *Verizon v. FCC*, 740 F.3d at 623, 650–51 (D.C. Cir. 2014).

21. Cf. Armin Grunwald, *Shaping the Present by Creating and Reflecting Futures*, in SOCIO-TECHNICAL FUTURES SHAPING THE PRESENT, EMPIRICAL EXAMPLES AND ANALYTICAL CHALLENGES 18 (Andreas Losch et al. eds., 2019) (“[I]n the present time we create futures supporting us to shape the present.”).

22. See Sandoval & Lanthier, *supra* note 18, at 6; JONATHAN NUECHTERLEIN & PHIL WEISER, DIGITAL CROSSROADS, TELECOMMUNICATIONS LAW AND POLICY IN THE INTERNET AGE 295 (2d ed. 2013).

23. Sandoval & Lanthier, *supra* note 18, at 6; Sandoval, *Net Neutrality Repeal Rips Holes in the Public Safety Net*, *supra* note 9, at 955 (“[T]he Internet and telephone networks[] rests on a distributed model of universal service that recognizes that society is better off when everyone has access to communications networks.”).

24. Sandoval & Lanthier, *supra* note 18, at 6 (citing *Texas Alarm & Signal Ass’n v. Pub. Util. Comm’n*, 603 S.W.2d 766, 770 (Tex. 1908); *Pub. Utility Comm’n of Texas v. AT & T Commc’ns of the Sw.*, 777 S.W.2d 363, 372–73 (Tex. 1989)).

During the COVID-19 pandemic, bandwidth-intensive, latency-sensitive internet use became an entry ticket for education, work, internet-based health services, and many other sectors. This tectonic shift underscores the imperative of ROBIN access and digital equity. “Digital equity refers to whether people can access and effectively use the technology necessary to participate in modern society.”²⁵ ROBIN access “is vital for the success of our communities, and it will become even more important as technology continues to advance and services continue to migrate online.”²⁶

Prior to the COVID-19 pandemic, in December 2019, videoconferencing platform Zoom had ten million daily participants; by October 2020, Zoom’s daily user base had grown to 300 million.²⁷ The 2021 Student Home Internet Connectivity study of several K-12 schools operating during the pandemic found that over “85 percent of network traffic in remote learning is used for video, which requires sufficient upload and download speeds. This increasingly popular learning trend is expected to continue for the foreseeable future.”²⁸ As COVID-19 infections challenge plans to return to school and work in person,²⁹ for the foreseeable future, use of video and other bandwidth-heavy and latency-sensitive platforms will likely continue as a major component of education, work, and other vital activities.

Millions of Americans, including more than 74 million prepaid mobile internet plan users in the United States³⁰ face ISP-imposed slowdowns, not readily discernable from ISP contracts. ISPs may limit subscriber internet use,

25. Zack Quaintance, *The Quest For Digital Equity: A Look at the Evolution of the Challenge to Ensure Advances in Technology Bring Benefits to Everyone*, GOV’T TECH. (Mar. 2018), <https://www.govtech.com/civic/the-quest-for-digital-equity.html>.

26. *Id.*

27. Rauf Arif, *In The Post COVID-19 World, Zoom is Here to Stay*, FORBES (Feb. 26, 2021, 1:18pm EST), <https://www.forbes.com/sites/raufarif/2021/02/26/in-the-post-covid-19-world-zoom-is-here-to-stay/?sh=50ab8b6d55b5>.

28. *Student Home Internet Connectivity Study*, COSN, <https://www.cosn.org/digitalequity> (last visited May 24, 2021).

29. See, e.g., Kristal Dixon, *Cobb School’s Fifth Grade Class Goes Virtual Due to Rise in Covid Cases*, ATLANTA J. CONST. (Aug. 11, 2021), <https://www.ajc.com/news/atlanta-news/cobb-schools-fifth-grade-class-goes-virtual-due-to-rise-in-covid-cases/4wpbplir55e2fgh6bo5vj3u54m/> (returning to virtual learning after 185 covid cases within the first few days of in-person school); Scott Jashick, *Delta Variant Raises Questions as Campuses Start Semester*, INSIDE HIGHER EDUC. (Aug. 16, 2021), <https://www.insidehighered.com/news/2021/08/16/delta-variant-raises-questions-colleges-about-reopening-plans> (reporting on colleges from Texas to California starting the fall semester online for two to six weeks due to increasing Covid-19 delta variant transmission); *Covid-19 Weekly Review*, CTRS. DISEASE CONTROL & PREVENTION (Aug. 13, 2021), <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html>.

30. Mark Lowenstein, *Lowenstein: What’s the Roadmap for Prepaid in the United States?*, FIERCE WIRELESS (Nov. 19, 2020, 12:36pm), <https://www.fiercewireless.com/wireless/lowenstein-what-s-roadmap-for-prepaid-united-states>.

constraining educational, health, and economic opportunities. As discussed in Section V.A, ISP software-imposed throttling to 2G speeds renders digital education, telehealth, work, and contemporary internet-based resources inaccessible for hours, days, or weeks at a time. ISP throttling harms internet users, critical infrastructure sectors and services, and our common future.

“Organizations socially construct risk and the choices available to increase or mitigate risk.”³¹ Regulation and regulatory enforcement allowing ISP throttling policies to persist—even when we know that doing so makes internet-based services inaccessible—creates risks to educational, health, and public safety service delivery. It wedges throttled consumers into the digital divide. Throttling creates risks to individuals, families, neighborhoods, and society, while undermining equity.

Throttling may affect a range of Americans including people who pay extra for high-end, unlimited internet access plans to support their work but are nonetheless throttled after their ISP deems them an excessive user.³² “Comcast announced in June 2018 that, ‘it no longer needs to throttle speeds for heavy internet users, ending a network-management technique it has been using since 2008.’”³³ Despite that policy change, Comcast emphasized it, “reserve[s] the right to implement a new congestion management system if necessary in the performance of reasonable network management and to maintain a good broadband internet access service experience for our customers, and will provide updates here as well as other locations if a new system is implemented.”³⁴

31. Catherine J.K. Sandoval, Pat Cain, Steve Diamond, Jean Love, Allen S. Hammond, Stephen E. Smith & Solmaz Nabipour, M.D., *Legal Education During the COVID-19 Pandemic: Put Health, Safety and Equity First*, 61 SANTA CLARA L. REV. 367, 454 (2020) (citing CINDY L. CALDWELL, SAFETY CULTURE AND HIGH-RISK ENVIRONMENTS: A LEADERSHIP PERSPECTIVE 4 (2018) (“The organization socially constructs a view that the essence of safety is to prevent individuals from committing errors.”); CHARLES PERROW, NORMAL ACCIDENTS: LIVING WITH HIGH-RISK TECHNOLOGY 372 (1984)).

32. Brodtkin, *supra* note 4 (stating Cox warned its customer who “pays \$150 a month, including \$100 for 1GBPs [one gigabit per second] download speeds and 35mbps [megabits per second] upload speeds, and another \$50 for ‘unlimited data’ so that he can go over Cox’s 1tb [terabyte] data cap, that he was engaging in ‘excessive use’ through his internet uploads, conducted primarily between 1:00 a.m. and 8:00 a.m., and that he and his neighborhood would have their data slowed”).

33. Sandoval, *Cybersecurity Paradigm Shift: The Risks of Net Neutrality Repeal to Energy Reliability, Public Safety, and Climate Change Solutions*, *supra* note 9, at 171; Liam Tung, *Comcast: We’ve Stopped Throttling Speeds for Heavy Internet Users, For Now*, ZDNET (June 14, 2018), <https://www.zdnet.com/article/comcast-weve-stopped-throttling-speeds-for-heavy-internet-users-for-now/> [<https://perma.cc/V3PH-EQTS>].

34. Tung, *supra* note 33.

Throttling is commonly found in contract terms for prepaid wireless internet service, making many smartphone-only internet users particularly vulnerable to this practice. “Reliance on smartphones for online access is especially common among younger adults, lower-income Americans and those with a high school education or less.”³⁵ Latinx and African Americans are more dependent on wireless internet service than Whites. Pew Research’s 2021 survey found “Black and Hispanic adults in the United States remain less likely than White adults to say they own a traditional computer or have high-speed internet at home,” but found “no statistically significant racial and ethnic differences when it comes to smartphone or tablet ownership.”³⁶

Deutsche Bank Research’s 2020 digital divide study found that “[d]ue to the structural and infrastructural inequities, Blacks and Hispanics are 10 years behind Whites in levels of broadband access and almost 4 times more Blacks have poor Tech connectivity than Whites.”³⁷ That study predicted that as a result of the digital divide, “76% of Blacks and 62% of Hispanics could get shut out or be under-prepared for 86% of jobs in the US by 2045. If this digital racial gap is not addressed, in one generation alone, digitization could render the country’s minorities into an unemployment abyss.”³⁸ ISP throttling is an underrecognized driver of poor broadband access and the digital divide that foments inequity and societal risk.

Deutsche Bank’s study highlights the lock-in effects of the digital divide and its likely effects on job access and the economy. Daria Roithmayr observed that “given the nature of the digital divide,” networks that “make use of the Internet to provide job referral networks targeting minority employees...are not likely to provide a critical density for job referral networks in segregated neighborhoods” as the legacy of residential segregation creates a “lock-in” effect that affects access to other resources.³⁹ Elizabeth Kennedy emphasized the importance of looking upstream at the

35. *Internet/Broadband Fact Sheet*, PEW RSCH. CTR. (Apr. 7, 2021), <https://www.pewresearch.org/internet/fact-sheet/internet-broadband/>.

36. Sara Atske & Andrew Perrin, *Home broadband adoption, computer ownership vary by race, ethnicity in the U.S.*, PEW RSCH. CTR. (July 16, 2021), <https://www.pewresearch.org/fact-tank/2021/07/16/home-broadband-adoption-computer-ownership-vary-by-race-ethnicity-in-the-u-s/>.

37. APJIT WALIA & SAI RAVINDRAN, DEUTSCHE BANK RSCH., AMERICA’S RACIAL GAP & BIG TECH’S CLOSING WINDOW 1 (Sept. 2, 2020), https://www.dbresearch.com/PROD/RPS_EN-PROD/America%27s_Racial_Gap_%26_Big_Tech%27s_Closing_Window/RPS_EN_DOC_VIEW.calias?rwnode=PROD0000000000464258&ProdCollection=PROD0000000000511664.

38. *Id.*

39. Daria Roithmayr, *Locked in Segregation*, 12 Va. J. Soc. Pol’y & L. 197 (2004).

causes of inequity.⁴⁰ It is not sufficient to count those stuck in the digital divide. We must identify its causes and practices that create a digital ceiling.

Adam J. Banks's observations about the digital divide can be applied to the FCC's failure to recognize ISP throttling to 2G speeds as a net neutrality problem. Banks wrote:

Thus, the Digital Divide involves both contest and silence: debate over whether there is a Divide at all, waged by politicians, foundations, and business interests now; debate over whether race is a factor in whatever problems in technology access might exist⁴¹

Assumptions about what constitutes digital technology access will “guide legal, corporate, and educational policies that can trap Black people into roles as passive consumers of technologies rather than producers and partners.”⁴² This process, Banks observed, can lead to “electronic invisibility and economic, educational, and political injustice.”⁴³

Deutsche Bank reported that “[a]lmost every professor we spoke to cited the digital divide during childhood as the cause of so many societal imbalances in today's digital age.”⁴⁴ That divide, baked in at childhood and persisting for many adults, fosters an inequitable future that undermines the economy, economic opportunity, and workforce preparedness.

Digital “inequity is spread across three spheres—access to Tech, Tech Training and Hardware.”⁴⁵ Access to Tech should be defined to include access to robust networks to enable high-speed internet functionality. Deutsche Bank proposes a “5-year program that seeks to reduce the gaps in the three areas targeting the underprivileged households among the Black and Hispanic communities [which] would cost approximately \$15BN,” a level the bank notes is “under 1% of the increase in the \$2 Trillion market cap the Big Tech 5 have had since Covid.”⁴⁶

This Article identifies regulatory law and policy as a source of inequity. Internet regulation produces or sustains equity or inequity. Internet governance, including regulation of disclosure policies, constitutes structural

40. Elizabeth J. Kennedy, *Desert in the Deluge: Using Data to Drive Racial Equity*, 69 CATH. U. L. REV. 23, 24 (2020).

41. ADAM J. BANKS, RACE, RHETORIC, AND TECHNOLOGY, SEARCHING FOR HIGHER GROUND 31 (2005).

42. *Id.*

43. *Id.*

44. WALIA & RAVINDRAN, *supra* note 37, at 6.

45. *Id.* at 7.

46. *Id.* at 7–8.

investments (or disinvestments) in community safety, resiliency, equity, and sustainable futures.⁴⁷ Failure to safeguard network, hardware, software, and training investments with laws and regulations that protect internet openness and safeguard consumers undercuts infrastructure access and effectiveness. Legal infrastructure is needed to protect investments in physical, software, and social capital infrastructure.

“Futures do not arise of their own accord. Techno-visionary futures are social constructs.”⁴⁸ Jan van Dijk observes that the digital divide is not a technical issue, “it is more of a social problem.”⁴⁹ Likewise, lack of ROBIN access is a socio-legal issue, the technical aspects of which are obscured through ISP nondisclosure of network management practices and their technical rationale, if any. Socio-legal decision-making constructs risks borne by the citizenry.⁵⁰ The limited legal imagination of regulators and lawmakers has left millions struggling to get or stay connected to the internet.⁵¹

In *Connect the Whole Community; Leadership Gaps Drive Disaster Vulnerability and the Digital Divide*, Patrick Lanthier and I discuss the dangerous conditions created by designing programs for an imagined community of highly connected citizens.⁵² During the 2017 flood that led to the evacuation of more than 14,000 people in San Jose, California when the Anderson dam overtopped, officials sent warnings to “an imagined community, highly connected to the [i]nternet, and capable of filtering warnings from the detritus of Twitter feeds, Facebook posts, and Nextdoor notices. In the process, officials failed to inform the community they served of the coming danger.”⁵³

Recognizing the difference between the real and imagined community is the first step to reforming program and regulatory design. Many utility and communications programs are designed for homeowners with little thought to the needs of renters. Many renters, as well as some homeowners, may not be able to install wired internet access due to landlord restrictions, living in households with multiple families, or living in temporary residences.

During my service as a California Public Utilities Commission (CPUC) Commissioner, I worked to reform energy efficiency programs to serve the

47. Sandoval & Lanthier, *supra* note 18, at 17.

48. Grunwald, *supra* note 21 at 24.

49. VAN DIJK, *supra* note 16, at 3.

50. Castells, *supra* note 16, at 7.

51. Cf. ELIZABETH FISHER, ENVIRONMENTAL LAW: A VERY SHORT INTRODUCTION 1 (2017) (“Legal imagination is needed to develop law to respond to a world of multiple interconnected parties, scientific uncertainty, and socio-political conflict.”).

52. See Sandoval & Lanthier, *supra* note 18, at 1, 6.

53. *Id.*

needs of renters and low-income Californians.⁵⁴ Such reforms pay additional dividends for internet-enabled energy demand response programs and communications-enabled initiatives.⁵⁵

The Supreme Court in *F.E.R.C. v. Electric Power Supply Ass'n.* observed that demand response “pays consumers for commitments to curtail their use of power, so as to curb wholesale rates and prevent grid breakdowns.”⁵⁶ California faced uncertainty about the sufficiency of its energy resources during mid-late 2021 while facing a historic drought.⁵⁷ Internet-enabled resources (such as connected thermostats) can reduce energy demand, forestall blackouts, and protect public safety.⁵⁸ Internet access at homes and businesses is crucial to the demand response programs that send signals to home or residential Wi-Fi to reduce air conditioning or other energy uses to stave off blackouts and protect public safety.⁵⁹

The FCC’s 2015 Open Internet Order (2015 OIO)—which adopted net neutrality rules that prohibited blocking, throttling, and paid priority—cited, as an example of public safety reasons to impose those regulations on ISPs, the comments I filed when I served as a CPUC Commissioner about the importance of a neutral and open internet to energy safety and reliability.⁶⁰ Rules protecting open internet access created confidence that led my CPUC colleagues and I to invest in ICT-enabled energy and water efficiency

54. CPUC D.16-11-022, Decision On Large Investor-Owned Utilities’ California Alternate Rates For Energy (CARE) And Energy Savings Assistance (ESA) Program Applications (Application 14-11-009), at 71, 120, 155, 161, 199, 256–57 (Nov. 10, 2016) (adopting Alternate Proposed Decision of Commissioner Catherine J.K. Sandoval).

55. Sandoval, *Net Neutrality Powers Energy and Forestalls Climate Change*, *supra* note 9, at 77.

56. *FERC v. Elec. Power Supply Ass'n*, 577 U.S. 260, 270 (2016), as revised (Jan. 28, 2016).

57. Joint Statement From the CPUC President Marybel Batjer, CEC Chair David Hochschild, and California ISO CEO Elliot Mainzer on Decision to Procure Additional Energy Resources for Summer, CAL. INDEP. SYS. OPERATOR (July 1, 2021), <https://www.aiso.com/Documents/CapacityProcurementMechanismSignificantEvent-JointStatementandLetter.pdf> (“As a result of these unprecedented climate change-driven heat events, which are occurring throughout the West in combination with drought conditions that reduce hydroelectric capacity, California is using all available tools to increase electricity reliability this summer.”).

58. Sandoval, *Net Neutrality Powers Energy and Forestalls Climate Change*, *supra* note 9, at 38–39 (citing In the Matter of Protecting & Promoting the Open Internet, 30 FCC Rcd. 5601, 5655 n. 291 (2015) [hereinafter FCC, 2015 Open Internet Order or 2015 OIO] (citing Catherine J.K. Sandoval, Commissioner, Cal. Pub. Util. Commission, Comment Letter on Protecting and Promoting the Open Internet, at 2 (Oct. 14, 2014) [hereinafter Commissioner Sandoval, *Ex Parte Letter*]).

59. Sandoval, *Net Neutrality Powers Energy and Forestalls Climate Change*, *supra* note 9, at 38–39.

60. FCC, 2015 Open Internet Order, *supra* note 58, at 5607, ¶¶ 5, 40.

programs and to expand those programs to low-income Californians including renters.⁶¹

Regulators must respond when programs and regulations are not meeting community needs. The Lifeline program was founded—first in California in 1984, then by the federal government in 1985—to support access to telephone services and thereby spur economic opportunity and protect public safety.⁶² California’s LifeLine program provides state financial support for telephone and internet access, complementing the federal Lifeline program administered by the FCC.

As the Assigned Commissioner for California’s LifeLine program, I held hearings that found Californians using wireless phones through the federal Lifeline program were running out of voice minutes, mostly by waiting on hold for social services, due to the lack of any FCC Lifeline minimum minutes standard then-existing.⁶³ My colleagues and I reformed California’s state LifeLine program in 2014, imposing minimum minutes standards for voice service to meet contemporary communications needs and opening the program to mobile platforms and internet use.⁶⁴ In 2016, the FCC followed California’s model by adopting minimum standards for voice minutes for the federal Lifeline program and focused its federal lifeline support on internet access.⁶⁵

61. Sandoval, *Net Neutrality Powers Energy and Forestalls Climate Change*, *supra* note 9, at 38–39.

62. CPUC, Resolution T-17366, Modifications To The California Lifeline Program Rules—General Order 153 - In Compliance with the Federal Communications Commission’s Lifeline/Link-Up Reform Order (FCC 12–11) (July 13, 2012), https://docs.cpuc.ca.gov/word_pdf/FINAL_RESOLUTION/170652.pdf (“[I]n 1984, the CPUC, established the Universal LifeLine Telephone Service Program in Decision (D.) 84-04-053.”); Cal. Pub. Utils. Code §§ 871–78 (introduced in 1983 as AB 1348, codified in 1987 as the Moore Universal Telephone Service Act); MTS and WATS Market Structure, and Amendment of Parts 67 & 69 of the Commission’s Rules and Establishment of a Joint Board, Report and Order, 50 Fed. Reg. 939 (1985) (implementing a low-income support program in 1985, after the divestiture of AT&T, that required carriers to offer discounted service to qualifying low-income consumers); Telecommunications Act of 1996, 47 U.S.C. § 224 (c)(1) (codifying “the commitment to advancing the availability of telecommunications services to low-income consumers and established principles upon which the Commission ‘shall base policies for the preservation and advancement of universal service.’”).

63. CPUC D. 14-01-036, Decision Adopting Revisions To Modernize And Expand The California Lifeline Program (Rulemaking 11-03-013) 56 (Jan. 16, 2014).

64. *Id.* at 71–73.

65. In the Matter of Lifeline & Link Up Reform & Modernization, 31 FCC Rcd. 3962 (2016) (“We also establish minimum service standards for broadband and mobile voice services to ensure those services meet the needs of the consumers, and we recognize and allow an exception in areas where fixed broadband providers do not meet the minimum standards.”).

The shortfall in voice service for consumers when the federal Lifeline program did not require carriers to provide any minimum level of voice minutes parallels the internet access deficit that happens to consumers throttled by their ISP. As discussed in Section V.B, throttling may occur after different levels of data consumption, depending on the plan and the ISP. Many throttled consumers are left without meaningful internet service, often for hours, days, or weeks at a time.⁶⁶ Throttling may leave consumers unable to conduct a telemedicine appointment, attend class or work via videoconferencing, or even to use mapping applications.

Switching to wired internet is not an option for many renters, those who share housing, or those who have insecure housing or poor credit. Neither are wired internet plans free from ISP throttling practices as some wired ISP terms of service claim the right to slow subscribers the ISP deems “excessive users,” even when “excessive” use is not clearly defined.⁶⁷ Some ISPs even throttle the neighbors of those it deems excessive users.⁶⁸

Analyzing ISP throttling practices and regulation is critical to ensure that the Emergency Broadband Benefit Program (EBB) program Congress authorized during the COVID-19 pandemic, adequately supports low-income consumer internet communications needs.⁶⁹ The FCC’s 2021 EBB order declined to impose any minimum standards and relied on ISP disclosures to support consumer choice.⁷⁰

Recognizing ISP throttling as a threat to internet openness is important to assessing America’s success in providing access to Advanced Telecommunications Service, as required by section 706 of the Communications Act.⁷¹ The FCC should examine the sufficiency of ISP disclosures of throttling practices and ensure that the EBB, Lifeline, and other FCC universal service and broadband programs meet contemporary communications needs. Addressing inequitable internet access, regulation, and ISP policies that undermine access is necessary to fulfill the FCC’s

66. See *supra* note 12 and *infra* Section V.B.

67. Brodtkin, *supra* note 4.

68. *Id.*

69. Consolidated Appropriations Act, 2021, Pub. L. No. 116-260, div. N, tit. IX, § 904(i), 134 Stat. 1182 (2020), <https://www.congress.gov/bill/116th-congress/house-bill/133/text> [hereinafter *Consolidated Appropriations Act*].

70. In the Matter of Emergency Broadband Benefit Program, WC Docket No. 20-445, Report and Order, ¶ 73 (Feb. 25, 2021), <https://docs.fcc.gov/public/attachments/FCC-21-29A1.pdf> [hereinafter, FCC, *EBB Order*].

71. See 47 U.S.C. § 1302(b) (requiring the FCC to conduct a notice of inquiry concerning the availability of advanced telecommunications capability to all Americans).

responsibility to promote access to advanced services, protect public safety, and serve all Americans.⁷²

III. LEGAL FRAMEWORK FOR FCC REGULATION TO PROMOTE SERVICE TO ALL AMERICANS IN THE PUBLIC INTEREST

A. PROMOTING WIRELESS AND WIRELINE SERVICE TO ALL AMERICANS WITHOUT DISCRIMINATION

The Communications Act requires the FCC to make wireless and wireline communications available to all of the people of the United States, without discrimination.⁷³ The Telecommunications Act of 1996 ('96 Act) added the requirement that the FCC carry out its mission “without discrimination on the basis of race, color, religion, national origin, or sex.”⁷⁴ David Honig, Co-Founder and former President and Executive Director of the Minority Media and Telecommunications Council, emphasized that the '96's Act's “non-discrimination provision is not self-executing.”⁷⁵

Federal statute mandates the FCC to regulate to promote “safety of life and property through the use of wire and radio communication.”⁷⁶ This requirement applies to the FCC's service to *all* the people of the United States. *Mozilla v. FCC* emphasized that “[w]hen, as here, ‘Congress has given an agency the responsibility to regulate a market such as the telecommunications industry that it has repeatedly deemed important to protecting public safety,’... the agency's decisions ‘must take into account its duty to protect the public.’”⁷⁷ The “Commission is ‘required to consider public safety by . . . its enabling act.’”⁷⁸

72. *See id.*; 47 U.S.C. §§ 151, 1302(b).

73. *See* 47 U.S.C. § 151.

74. 110 Stat. 86, Sec. 104, Nondiscrimination Principle, Pub. Law 104, Feb. 8, 1996.

75. David Honig, *How the FCC Suppressed Minority Broadcast Ownership, and How the FCC Can Undo the Damage It Caused*, 12 S. J. POL'Y & JUST. 44, 47 (2018).

76. *See* *Nuvio Corp. v. F.C.C.*, 473 F.3d 302, 311 (D.C. Cir. 2006) (discussing the FCC's statutory duty to promote public safety); *Mozilla v. FCC*, 940 F.3d 1, 94–97 (D.C. Cir. 2019) (citing Professor and former CPUC Commissioner Sandoval's comments about the internet's role in public safety, energy reliability and safety, natural gas leak detection, and critical infrastructure protection, in addition to CPUC and County of Santa Clara's comments, to remand the FCC's net neutrality repeal order to consider public safety issues); *see also* Wireless Communication and Public Safety Act of 1999, 47 U.S.C. § 615 (requiring the FCC to promote safety through its regulation of wireless communications).

77. *Mozilla v. FCC*, 940 F.3d 1, 7–8 (citing *Nuvio*, 473 F.3d at 307).

78. *Id.* at 59.

The radio spectrum, including that used for wireless and wireline communication, belongs to the people of the United States. The Communications Act provides, “[i]t is the purpose of this [Act], among other things, to maintain the control of the United States over all the channels of radio transmission.”⁷⁹ FCC licenses “provide for the use of such channels, but not the ownership thereof, by persons for limited periods of time, under licenses granted by Federal authority, and no such license shall be construed to create any right, beyond the terms, conditions, and periods of the license.”⁸⁰

The communications sector is a part of the economy designated as “Critical Infrastructure.”⁸¹ The Critical Infrastructure Protection Act (CIPA) of 2001 defines critical infrastructure as those “systems and assets, whether physical or virtual, so vital to the United States that the incapacity or destruction of such systems and assets would have a debilitating impact on security, national economic security, national public health or safety, or any combination of those matters.”⁸² CIPA “defines critical infrastructure *not* with reference to the identity of the target, but by the consequences of an attack on it.”⁸³

In 2013, President Obama issued the Presidential Policy Directive-Critical Infrastructure Security and Resilience (PPD-21), which designated 16 sectors as critical infrastructure, including the Communications Sector.⁸⁴ PPD-21 identifies “energy and communications systems as uniquely critical due to the enabling functions they provide across all critical infrastructure sectors.”⁸⁵ “Energy and communications systems are key drivers for the U.S. economy, democracy, and national security, underlying the operations of nearly all businesses, public safety organizations, healthcare providers, education, and government.”⁸⁶

Pursuant to CIPA, the Communications Sector Specific plan “outlines how government and private sector participants in the critical infrastructure

79. 47 U.S.C. § 301.

80. *Id.*

81. *Presidential Policy Directive—Critical Infrastructure Security And Resilience (Ppd-21)*, WHITE HOUSE (Feb. 12, 2013), <https://obamawhitehouse.archives.gov/the-press-office/2013/02/12/presidential-policy-directive-critical-infrastructure-security-and-resil>; *Communications Sector*, CYBERSECURITY & INFRASTRUCTURE SEC. AGENCY, <https://www.cisa.gov/communications-sector>.

82. Sandoval, *Net Neutrality Powers Energy and Forestalls Climate Change*, *supra* note 9, at 8 (citing 42 U.S.C. § 5195c(e)).

83. *Id.*

84. WHITE HOUSE, *supra* note 81.

85. *Id.*

86. Sandoval, *Net Neutrality Powers Energy and Forestalls Climate Change*, *supra* note 9, at 11.

community work together to manage risks and achieve security and resilience outcomes.”⁸⁷ The Communications Sector plan recognizes that changes in the communications industry—“including mobile broadband, cloud computing, the Internet of Things (IoT),” “software-defined networks (SDNs),” and widespread smartphone and tablet computer adoption—have created “enormous demand for mobile broadband communications,” as well as increased “the requirement for improved sector security and resilience.”⁸⁸

The Communications Sector plan is network-focused rather than consumer-focused. The Communications Sector’s classification as Critical Infrastructure underscores the importance of internet network access to national economic security, national public health, and safety. Addressing throttling, industry practices, and regulations that limit consumer internet access is critical to achieving CISA’s goals.

The Telecommunications Act of 1996, section 706(b), requires the Commission to annually “initiate a notice of inquiry concerning the availability of advanced telecommunications capability to all Americans (including, in particular, elementary and secondary schools and classrooms)”⁸⁹ Through this annual inquiry, the FCC must “determine whether advanced telecommunications capability is being deployed to all Americans in a reasonable and timely fashion.”⁹⁰ “If that determination is negative, the Commission ‘shall take immediate action to accelerate deployment of such capability by removing barriers to infrastructure investment and by promoting competition in the telecommunications market.’”⁹¹

The FCC is also required, by RAY BAUM’S Act of 2018 to issue a biennial Communications Marketplace Report assessing “the state of deployment of communications capability, including advanced

87. *National Infrastructure Protection Plan*, CYBERSECURITY & INFRASTRUCTURE SEC. AGENCY, <https://www.cisa.gov/national-infrastructure-protection-plan>; U.S. DEPT HOMELAND SEC., COMMUNICATIONS SECTOR-SPECIFIC PLAN AN ANNEX TO THE NIPP 2013 1 (2015) [hereinafter *Communications Sector-Specific Plan*], <https://www.cisa.gov/sites/default/files/publications/nipp-ssp-communications-2015-508.pdf>.

88. *Id.* at 5.

89. 47 U.S.C. § 1302(b).

90. *Id.*

91. FCC, FCC 20-50, 2020 BROADBAND DEPLOYMENT REPORT ¶ 4 (2020) (citing Consolidated Appropriations Act, 2018, Pub. L. No. 115-141, Div. P—RAY BAUM’S Act of 2018, §§ 401–402, 132 Stat. 348, 1087–90 (2018) (RAY BAUM’S Act of 2018); 47 U.S.C. § 163(b)(2) (added 2018)), <https://www.fcc.gov/reports-research/reports/broadband-progress-reports/2020-broadband-deployment-report> [hereinafter BROADBAND DEPLOYMENT REPORT].

telecommunications capability.”⁹² The FCC issues both the Broadband Deployment Report required under section 706(b) to assess the state of advanced service deployment and a Communications Marketplace Report in compliance with the RAY BAUM Act.⁹³

This Article recommends the FCC examine restrictive throttling practices in its Communications Marketplace Report and section 706 reports on Advanced Communications. That analysis should consider whether ISPs should be required to publicly disclose throttling practice information, even if the ISP characterizes that conduct as network management. Information about the number of Americans subject to throttling, the extent and duration of such slow-downs, and the neighborhoods where throttling is practiced can inform FCC regulation and assessments to promote internet access and protect public safety.

B. PROMOTING ADVANCED TELECOMMUNICATIONS SERVICES AND THE OPEN INTERNET’S VIRTUOUS CIRCLE OF INNOVATION

In addition to requiring the FCC to assess the status of advanced telecommunications service, section 706(b) empowers the FCC to “promulgate rules governing broadband providers’ treatment of [i]nternet traffic.” The D.C. Circuit, in *Verizon v. FCC*, upheld the FCC’s 2010 Open Internet Order’s (2010 OIO) interpretation of section 706(b) as an affirmative grant of authority allowing the FCC to take action to promote internet access.⁹⁴ The FCC’s 2010 OIO reversed the FCC’s previous interpretation of section 706 as not authorizing such action.⁹⁵ Based on the FCC’s earlier interpretation of section 706 as not granting affirmative authority to adopt rule to promote internet access, the D.C. Circuit concluded in 2010 that the FCC did not have the jurisdictional basis to order sanctions against Comcast for its actions alleged to have interfered with internet traffic.⁹⁶

Verizon v. FCC recognized the FCC’s theory that regulatory action to protect internet openness was appropriate and authorized by section 706 to

92. *Id.* at 11; *see also* Communications Marketplace Report, 33 FCC Rcd. 12558, 12683–702, ¶¶ 236–64 (2018) [hereinafter *Communications Marketplace Report*].

93. *See, e.g.,* *Communications Marketplace Report*, *supra* note 92.

94. *Verizon v. FCC*, 740 F.3d 623, 627 (D.C. Cir. 2014).

95. *In re Preserving the Open Internet*, 25 F.C.C.R. 17905 (2010) (reversing the FCC’s prior interpretation of section 706(b) of the Telecom Act and recognizing that statute as an affirmative grant of authority to promote advanced internet access) [hereinafter *FCC, 2010 Open Internet Order* or *2010 OIO*].

96. *Cf. Comcast Corp. v. FCC*, 600 F.3d 642 (D.C. Cir. 2010) (holding the FCC failed to cite any statutory authority to justify its order compelling a broadband provider to adhere to open network management practices, based on the then-existing interpretation of section 706(b) as not granting the FCC authority to take action to promote advanced internet access).

“preserve and facilitate the ‘virtuous circle’ of innovation that has driven the explosive growth of the [i]nternet”⁹⁷ That virtuous circle of internet innovation turns on internet openness that “spurs investment and development by edge [content] providers, which leads to increased end-user demand for broadband access.”⁹⁸ Innovative internet content and platforms lead “to increased investment in broadband” infrastructure.⁹⁹

The FCC adopted rules to protect internet openness because they were concerned that if broadband providers were to disrupt this “virtuous circle” by “[r]estricting edge providers’ ability to reach end users and limiting end users’ ability to choose which edge providers to patronize,” providers would “reduce the rate of innovation at the edge and, in turn, the likely rate of improvements to network infrastructure.”¹⁰⁰ The D.C. Circuit concurred with the FCC’s judgment that section 706(b) of the ’96 Act vests it with “affirmative authority to enact measures encouraging the deployment of broadband infrastructure.”¹⁰¹

The FCC adopted rules in its 2010 OIO under the FCC’s Title I authority conferred by the Communications Act of 1934. The Communications Act founded the FCC to make wireless and wireline communications available to all of the people of the United States, to protect safety of life and property, and for national defense.¹⁰² The FCC “justified its [2010 Open Internet Order] as an exercise of what courts term its ‘ancillary jurisdiction,’ a power that flows from the broad language of Communications Act (47 U.S.C. § 154(i)).”¹⁰³ That statute authorizes the Commission to “perform any and all acts, make such rules and regulations, and issue such orders, not inconsistent with this chapter, as may be necessary in the execution of its functions.”¹⁰⁴

The Supreme Court has held that the FCC may exercise such ancillary jurisdiction where two conditions are met: “(1) the Commission’s general jurisdictional grant under Title I covers the regulated subject and (2) the regulations are reasonably ancillary to the Commission’s effective performance of its statutorily mandated responsibilities.”¹⁰⁵ *Verizon v. FCC*

97. *Verizon*, 740 F.3d at 627.

98. *Id.* at 634.

99. *Id.*

100. FCC, 2010 Open Internet Order, *supra* note 95, ¶ 14.

101. *Verizon*, 740 F.3d at 628.

102. 47 U.S.C. § 151, 154(i).

103. *Verizon*, 740 F.3d at 632.

104. *Id.*

105. *Id.* (citing *Am. Library Ass’n v. FCC*, 406 F.3d 689, 691–92); *see also* *United States v. Sw. Cable Co.*, 392 U.S. 157 (1968); *United States v. Midwest Video Corp.*, 406 U.S. 649 (1972); *FCC v. Midwest Video Corp.*, 440 U.S. 689 (1979).

determined that since the FCC had not yet interpreted section 706 of the Communications Act to grant it affirmative authority to take regulatory action to promote internet access, the ancillary jurisdiction test was not met to adopt regulations prohibiting ISPs from blocking and throttling.¹⁰⁶

Verizon v. FCC concluded that the FCC's requirements that broadband providers "serve all edge providers without 'unreasonable discrimination,'" "relegated [those providers], *pro tanto*, to common carrier status" under Title II of the Communications Act.¹⁰⁷ The D.C. Circuit determined that FCC rules against blocking and throttling could not be sustained based on the FCC's classification of ISPs under Title I in its 2010 OIO.¹⁰⁸

The 2010 OIO also adopted transparency rules requiring fixed and mobile broadband providers to "publicly disclose accurate information regarding the network management practices, performance, and commercial terms of [their] broadband [i]nternet access services."¹⁰⁹ The D.C. Circuit upheld the FCC's transparency and disclosure rules as severable from its rules prohibiting blocking and imposing antidiscrimination rules under Title I of the Communications Act.¹¹⁰

In response to these legal precedents and a robust record expressing concern about threats to internet openness, the FCC's 2015 OIO classified ISPs under Title II of the Communications Act and imposed rules prohibiting ISP blocking, throttling, and paid priority.¹¹¹ FCC Title II regulatory classification of ISPs satisfied the D.C. Circuit's concern about the FCC's authority to prohibit ISP blocking and throttling, regulations that treated ISPs like common carriers by subjecting them to nondiscrimination obligations.

The FCC's 2015 OIO prohibited ISPs, classified as common carriers, from engaging in throttling, including "the degrading of [i]nternet traffic

106. *Verizon*, 740 F.3d at 656.

107. *Verizon*, 740 F.3d at 624, 633, 658 (citing *Midwest Video II*, 440 U.S. 689, 700–01 (1979); *NARUC I v. FCC*, 525 F.2d 630, 642 (D.C. Cir. 1976)).

108. *Verizon*, 740 F.3d at 650–51 ("Thus, we must determine whether the requirements imposed by the Open Internet Order subject broadband providers to common carrier treatment. If they do, then given the manner in which the Commission has chosen to classify broadband providers [as information service providers under Title I], the regulations cannot stand.").

109. *Id.* at 659 (upholding the FCC's 2010 *Open Internet Order* transparency rules and reversing the rules against blocking and throttling as common carrier-type restrictions, not supported by the FCC's classification of ISPs as information service providers); FCC, 2010 *Open Internet Order*, *supra* note 95, ¶¶ 54 (transparency rules for fixed providers), 98 (transparency rules for mobile providers).

110. *Verizon*, 740 F.3d at 656.

111. FCC, 2015 *Open Internet Order*, *supra* note 58, at 5604.

based on source, destination, or content.”¹¹² Proscriptions against throttling were adopted to “avoid gamesmanship designed to avoid the no-blocking rule by, for example, rendering an application effectively, but not technically, unusable.”¹¹³

The FCC’s 2015 OIO also imposed a rule against unreasonable interference by ISPs, also known as the “general conduct rule.”¹¹⁴ The rule allowed complaints to be filed with the FCC against ISP policies and practices not clearly proscribed by the bright-line rules against blocking, throttling, and paid priority (receiving payment to prioritize some internet traffic).¹¹⁵

The 2015 OIO also required disclosure of network management practices and did not deem business justifications—as opposed to technical network management reasons—as reasonable network management.¹¹⁶ It modified the transparency rules adopted in the 2010 Order to require “specific notification to consumers that a ‘network practice’ is likely to significantly affect their use of the service.”¹¹⁷

The D.C. Circuit in *U.S. Telecom Ass’n v. FCC (USTA)* upheld the FCC’s decision to classify ISPs under Title II as a reasonable exercise of the FCC’s authority under section 706(b) to interpret an ambiguous statute.¹¹⁸ *USTA* found that the FCC’s decision to classify ISPs as Title II carriers and impose antidiscrimination rules was a result of reasoned decision-making due *Chevron* deference.¹¹⁹ *USTA* upheld the FCC’s 2015 Order, citing the FCC’s analysis that “convincingly detailed how broadband providers’ [gatekeeper] position in the market gives them the economic power to restrict edge-provider [content provider] traffic and charge for the services they furnish edge providers.”¹²⁰

After Presidential administrations changed in 2017, the FCC, in 2018, reversed course and classified ISPs under Title I, repealing the 2015 rules.

112. *Id.* at 5607, ¶ 17.

113. *Id.*

114. *See* *U.S. Telecom Ass’n v. FCC*, 825 F.3d 674, 733 (D.C. Cir. 2016) (citing FCC, *2015 Open Internet Order*, *supra* note 58, at 5659–60, ¶ 136); FCC, *2015 Open Internet Order*, *supra* note 58, at 5659–68, 5728–30, ¶¶ 135–53, 294–96.

115. *Id.* at 5728–29, 5885, App. A, Sec. 8.11 (imposing a no unreasonable interference/disadvantage standard to ensure that broadband providers do not engage in practices that threaten the internet’s open nature in other or novel ways).

116. *Id.* at 5700, ¶¶ 215–16.

117. *Id.* at 5609, ¶ 24.

118. *U.S. Telecom Ass’n*, 825 F.3d at 694.

119. *Id.* at 689, 697 (citing *Chevron U.S.A. Inc. v. Nat. Res. Def. Council, Inc.*, 467 U.S. 837 (1984)).

120. Sandoval, *Net Neutrality Repeal Rips Holes in the Public Safety Net*, *supra* note 9, at 988–89.

Citing a policy preference for light-touch regulation and reliance on the FTCA and antitrust law to protect consumers and competition, the FCC divested its authority to impose net neutrality rules against blocking, throttling, and paid priority or other nondiscrimination obligations.¹²¹

The 2018 “Restoring Internet Freedom” (RIF) Order required ISPs to disclose throttling defined as “[a]ny practice (other than reasonable network management elsewhere disclosed) that degrades or impairs access to lawful [i]nternet traffic on the basis of content, application, service, user, or use of a non-harmful device, including a description of what is throttled.”¹²² Under those rules, no disclosure would be required if the practice were “reasonable network management elsewhere disclosed.”¹²³ The FCC’s 2018 rules gave ISPs great latitude to determine what is reasonable network management and effectively limited FCC jurisdiction to determine what is reasonable.

The 2018 RIF Order modified the FCC’s transparency rules to require ISPs to disclose network management practices and other terms to inform consumer choice.¹²⁴ The FCC added to its rules in 47 C.F.R. § 8.1 a requirement for ISPs to disclose congestion management policies:

Any person providing broadband [i]nternet access service shall publicly disclose accurate information regarding the network management practices, performance characteristics, and commercial terms of its broadband [i]nternet access services sufficient to enable consumers to make informed choices regarding the purchase and use of such services and entrepreneurs and other small businesses to develop, market, and maintain [i]nternet offerings. Such disclosure shall be made via a publicly available, easily accessible website or through transmittal to the Commission.¹²⁵

121. FCC, *2018 RIF Order*, *supra* note 12, at 7852 (repealing FCC rules adopted in 2015 that prohibited ISPs from blocking, throttling, or engaging in paid prioritization of internet traffic except for limited reasonable network management justifications).

122. *Id.* ¶ 201 (requiring disclosure of throttling practices).

123. *Id.*

124. *Id.* (“Descriptions of congestion management practices, if any [are required]. These descriptions should include the types of traffic subject to the practices; the purposes served by the practices; the practices’ effects on end users’ experience; criteria used in practices, such as indicators of congestion that trigger a practice, including any usage limits triggering the practice, and the typical frequency of congestion; usage limits and the consequences of exceeding them; and references to engineering standards, where appropriate.”).

125. *Id.* ¶ 5 (amending 47 C.F.R. § 8.1).

Through that transparency rule, the FCC sought to “provide edge [content] providers with the information necessary to develop new content, applications, services, and devices that promote the virtuous cycle of investment and innovation.”¹²⁶ It also sought to “better enable end-user consumers to make informed choices about broadband services by providing them with timely information tailored more specifically to their needs”¹²⁷

As discussed in Section V.B., many ISPs fail to disclose the extent or consequences of throttling, such as the inability to access telemedicine, education, or work through videoconferencing, in addition to GPS navigation and other commonly used applications. This lack of disclosure should be examined as it appears inconsistent with the FCC transparency policy’s requirement that ISPs disclose the consequences of exceeding congestion management policies or limits.

The D.C. Circuit in *Mozilla v. FCC* upheld the FCC’s authority to reclassify ISPs under Title I but remanded the 2018 RIF Order for failure to consider the impact of net neutrality repeal on public safety, ISP access to utility poles, and Lifeline service for the poor—the latter two of which require common carrier classifications.¹²⁸ *Mozilla* cited my comments filed with the FCC in 2014 when I was a CPUC Commissioner, as well as those of Santa Clara County and the CPUC, as a warning about the dangers to public safety flowing from ISP interference with internet openness.¹²⁹ My comments highlighted many safety-related benefits of an open internet—safeguarded from blocking, throttling, and paid priority—exemplified by uses in the fields of energy, water, and communications. For example, “[i]nternet-enabled demand response transforms load reduction into a supply-side energy resource that manages energy during critical events, and forestalls the need to build fossil-fueled power plants.”¹³⁰

The D.C. Circuit chastised the FCC for its failure to analyze “the direct and specific comments by Santa Clara County, former California Public Utility Commissioner Sandoval, and others” that “repeatedly raised

126. *Id.* ¶ 162.

127. *Id.*

128. *Mozilla Corp. v. FCC*, 940 F.3d 1, 41 (D.C. Cir. 2019).

129. *Id.* at 61–62.

130. Catherine J.K. Sandoval, Commissioner, California Public Utilities Commission, Testimony and statement Before the Congressional Democratic Forum on Net Neutrality, Hosted by Congresswoman Doris O. Matsui, September 24, 2014, Written Statement of Testimony Submitted to the FCC record for GN Docket Nos., 14-28, and 10-127, Protecting and Promoting the Open Internet, Framework for Broadband Internet Services, 34–35, <https://ecfsapi.fcc.gov/file/60000972787.pdf> [hereinafter *Sandoval Net Neutrality September 2014 Testimony*].

substantial concerns about the Commission’s failure to undertake the statutorily mandated analysis of the 2018 Order’s effect on public safety.”¹³¹ The D.C. Circuit emphasized that since promoting public safety is a statutory duty under 47 U.S.C. § 151 and other statutes, the FCC acted arbitrarily and capriciously by failing to consider the link between net neutrality and public safety under the Administrative Procedure Act.¹³²

Santa Clara County’s comments in the Government Petitioner’s brief in *Mozilla v. FCC* emphasized the public safety harms caused by blocking, throttling, and paid priority. Santa Clara County expressed grave concern that, in July 2017, “Verizon slowed the Santa Clara Fire Protection District’s data when the District was fighting the Mendocino Complex fire—California’s largest fire,” as of that date.¹³³ “During this slowdown, Fire District personnel appealed to Verizon to stop the severe data slowdown for a device in active use to help coordinate fire resources.”¹³⁴

As Andrea Matwyshyn explained, the emails between Verizon and the Santa Clara County Fire Department showed

Verizon responded by informing them in the middle of this public safety emergency that [i]nternet access had been throttled because the department had purchased an ‘incorrect’ tier of service: the ‘unlimited’ plan the department had purchased was contractually subject to throttling in the sole discretion of Verizon. The exchange between Verizon and the firefighters was memorialized in a series of progressively more desperate, plaintive emails from the firefighters to Verizon.¹³⁵

“Even after they explained the gravity of the situation, the firefighters perceived the Verizon representative to be more concerned with attempting to upsell the department on a higher tier of service than assisting them during the crisis.”¹³⁶ “Verizon demanded that the Fire Department switch to a plan that costs \$2.00 a month more to stop the throttling, an unfathomable demand to a fire department using the [i]nternet during an active firefight. Fire Department personnel could not readily authorize additional payments

131. *Id.* 60–61; *see also* Brief for Government Petitioners at 23, *Mozilla v. FCC*, 940 F.3d 1 (2018) (No. 18-1051) (citing *Sandoval Net Neutrality September 2014 Testimony*, *supra* note 130, at 34–35).

132. *Sandoval, Net Neutrality Repeal Rips Holes in the Public Safety Net*, *supra* note 9 at 1017.

133. *Id.*

134. *Id.*

135. Andrea M. Matwyshyn, *Unavailable*, 81 U. PITT. L. REV. 349, 369 (2019).

136. *Sandoval, Net Neutrality Repeal Rips Holes in the Public Safety Net*, *supra* note 9 at 1017.

for the requested \$2.00 per month upcharge in light of government contracting rules.”¹³⁷

The “unlimited” data plan Verizon sold to the Santa Clara Fire Department did not sufficiently highlight the potential throttling that thwarted internet access necessary to coordinate resources during a fire fight. “Throttling means that the device that can normally act like a modern broadband internet connection is slowed to the point of acting more like an AOL dial up modem from 1995,” Santa Clara’s Fire Chief reported.¹³⁸ “Verizon’s service slowdown turned the [i]nternet calendar back to the dial-up days in the midst of a public safety emergency. Throttling left firefighters unable to use data connections that require more than dial-up speeds to acquire information and coordinate their firefighting response.”¹³⁹ Santa Clara County Fire Chief Anthony Bowden “asserted that the department ‘experienced throttling by its ISP, Verizon,’ namely that its ‘data rates had been reduced to 1/200, or less, than the previous speeds.’”¹⁴⁰ This throttling was dangerous because “[t]he [i]nternet has become an essential tool in providing fire and emergency response,”¹⁴¹ and “[m]odern firefighters rely on real-time geographic information system (GIS) mapping to monitor fires and coordinate emergency response, track information, and save lives.”¹⁴²

In response to Fire Chief Bowden, Verizon asserted, “ ‘This was a customer support mistake’ and not a net neutrality issue.”¹⁴³ Verizon’s response finds support from Jonathan E. Nuechterlein and Howard Shelanski who argued that advocates “sometimes use the term ‘throttling’ to describe the slower speeds that customers on tiered data plans sometimes experience after they have exceeded their monthly data allowances.”¹⁴⁴ They contend that this “practice has nothing to do with discriminating among content sources or preserving an open internet, and it has always been lawful, even under the now-repealed Title II regime.”¹⁴⁵ The Fire Department argued “Verizon’s throttling has everything to do with net neutrality—it shows that

137. *Id.*

138. *Id.* (citing Addendum to Brief for Government Petitioners at Appx. A, 11, *Mozilla Corp. v. FCC*, 940 F.3d 1 (2018) (No. 18-1051)).

139. *Id.*

140. Christopher Terry & Scott Memmel, *Harlem Shake Meets the Chevron Two Step: Net Neutrality Following Mozilla v. FCC*, 15 WASH. J. L. TECH. & ARTS 160, 162 (2020).

141. Matwyshyn, *supra* note 135, at 370.

142. Sandoval, *Net Neutrality Repeal Rips Holes in the Public Safety Net*, *supra* note 9, at 1045.

143. Matwyshyn, *supra* note 135, at 370.

144. Jonathan E. Nuechterlein & Howard Shelanski, *Building on What Works: An Analysis of U.S. Broadband Policy*, 73 FED. COMM. L.J. 219, 257 n. 82 (2021).

145. *Id.*

the ISPs will act in their economic interests, even at the expense of public safety.”¹⁴⁶

Verizon’s argument that throttling the Fire Department was not a net neutrality issue focused on whether the 2015 Open Internet rules would have applied to an enterprise customer such as the Santa Clara County fire department.¹⁴⁷ This argument misses the point. Verizon’s throttling of the fire department during an active fire fight is emblematic of the type of throttling untold numbers of customers experience when their ISP slows their use to dial-up speeds. As discussed in Section V.B., several ISPs throttle customers to a similar extent as Verizon’s deleterious throttling of the Santa Clara County Fire Department’s service.

The Santa Clara County Fire Department “throttling incident galvanized the California state legislature and citizenry, and net neutrality legislation passed shortly thereafter.”¹⁴⁸ I appreciated the opportunity to suggest to Senator Wiener and his staff language incorporated into the California Internet Consumer Protection and Net Neutrality Act of 2018, Senate Bill 822 (SB 822), section 1. The language “recogniz[es] the importance of the [i]nternet to critical infrastructure services, the economy, businesses, and other activities regulated by the state’s police power.”¹⁴⁹ Section 1 states that the act “is adopted pursuant to the police power inherent in the State of California to protect and promote the safety, life, public health, public convenience, general prosperity, and well-being of society, and the welfare of the state’s population and economy, that are increasingly dependent on an open and neutral [i]nternet.”¹⁵⁰

As enacted, SB 822 prohibits fixed internet service providers from engaging in blocking, throttling, and paid priority, among other practices. Title 15, section 3101(a), of the California Civil Code, prohibits “a fixed [i]nternet service provider” from “[i]mpairing or degrading lawful [i]nternet traffic on the basis of [i]nternet content, application, or service, or use of a nonharmful device, subject to reasonable network management.” This

146. Matwyshyn, *supra* note 135, at 370.

147. Jon Brodtkin, *Fire Dept. Rejects Verizon's "Customer Support Mistake" Excuse for Throttling*, ARS TECHNICA (Aug. 22, 2018), <https://arstechnica.com/tech-policy/2018/08/fire-dept-rejects-verizons-customer-support-mistake-excuse-for-throttling/>.

148. Matwyshyn, *supra* note 135, at 370; S.B. 822, 2018 Leg., Reg. Sess. (Cal. 2018), https://leginfo.ca.gov/faces/billNavClient.xhtml?bill_id=201720180SB822.

149. Sandoval, *Net Neutrality Powers Energy and Forestalls Climate Change*, *supra* note 19, at 21 (citing *Gonzales v. Oregon*, 546 U.S. 243 (2006) (providing that states have authority under the police power to “legislate with regard to protection of the lives, limbs health, comfort, and quiet of all persons”).

150. Cal. S.B. 822 § 1.

prohibition does not apply to wireless carriers or to service degradation agnostic to content, application, service, or device.

Section 3101(7)(A) prohibits unreasonable network management:

Unreasonably interfering with, or unreasonably disadvantaging, either an end user's ability to select, access, and use broadband Internet access service or the lawful Internet content, applications, services, or devices of the end user's choice, or an edge provider's ability to make lawful content, applications, services, or devices available to end users. Reasonable network management shall not be a violation of this paragraph.¹⁵¹

Whether throttling consumers to 2G speeds, or other types of throttling, is “reasonable network management” will likely be a subject for future litigation.

Several communications industry associations challenged SB 822's legality, arguing federal preemption prohibits California from adopting its own net neutrality rules due to the internet's interstate nature.¹⁵² Notably, *Mozilla v. FCC* vacated the FCC's 2018 RIF Order's attempt to preempt state ISP regulation as having no basis in statute or authority.¹⁵³ Thus, in *ACA CONNECTS*, amici supporting the state argued:

SB 822 protects Californians by filling the gap the FCC created. It preserves the ability of all Californians, including the most vulnerable, to fully participate—economically, socially, and politically—in everyday life. It ensures that ISPs cannot block, throttle, or distort [i]nternet content, and prohibits zero-rating and other paid-prioritization schemes so that all Californians have full access to lawful [i]nternet content and services—at lower prices.¹⁵⁴

The Ninth Circuit upheld the district court's denial of a preliminary injunction to block SB 822's enforcement, concluding, consistent with *Mozilla v. FCC*, that “by classifying broadband internet services as

151. *Id.*

152. Reply Brief of Plaintiffs-Appellants Broadband Provider Associations at 1, *ACA Connects v. Bonta*, No. 21-15430 (9th Cir. May 25, 2021), 2021 WL 2273765.

153. *Mozilla Corp. v. FCC*, 940 F.3d 1, 86 (D.C. Cir. 2019) (“At bottom, the Commission lacked the legal authority to categorically abolish all fifty States' statutorily conferred authority to regulate intrastate communications. For that reason, we vacate the Preemption Directive.”).

154. Brief of Amici Curiae Electronic Frontier Foundation, ACLU Foundation of Northern California, ACLU Foundation of Southern California, Access Humboldt, Benton Institute for Broadband & Society, Clean Money Campaign, Fight for the Future, Greenling Institute, Ifixit, Inc., Media Justice, National Hispanic Media Coalition, Oakland Privacy, Reddit, Inc., Turn - The Utility Reform Network, Writers Guild of America, West, Inc. in Support of Defendant-Appellee and Affirmance at 4, *ACA Connects v. Bonta*, No. 21-15430 (9th Cir. May 11, 2021), 2021 WL 2022131.

information services, the FCC no longer has the authority to regulate in the same manner that it had when these services were classified as telecommunications services.” The agency, therefore, cannot preempt state action, like SB-822, that protects net neutrality.”¹⁵⁵

SB 822 reflects the state’s concern with ISP actions that affect internet openness. Throttling was a primary motivator for SB 822’s adoption and is a practice affecting a range of consumers from mass-market internet users to institutional internet users such as a fire department. The issues leading up to SB 822 and its text highlight the safety and public interest issues at stake in ISP conduct and regulation.

The FCC concluded in October 2020, on remand from *Mozilla v. FCC*, that Title I classification for ISPs was unlikely to adversely affect public safety.¹⁵⁶ The FCC’s order prioritized the potential benefits of Title I classification over likely public safety risks. The 2020 RIF Order on remand determined that “even if there were some adverse impacts on public safety applications in particular cases—which we do not anticipate—the overwhelming benefits of Title I classification would still outweigh any potential harms.”¹⁵⁷ This conclusion and choice between classifying ISPs under Title I or Title II are likely to be revisited in future FCC net neutrality proceedings, which are outside of this Article’s scope.

Separate from net neutrality analysis, this Article recommends that the FCC evaluate ISP throttling and contract terms under its transparency rules still standing after the 2018 RIF Order and 2020 remand. That analysis should consider whether ISP throttling policy disclosures are sufficient to protect public safety, a value increasingly dependent on robust and open internet access.

C. CENTERING THE PUBLIC IN PUBLIC SAFETY

In its order on remand from *Mozilla v. FCC*, the FCC’s public safety definition excluded communications among members of the public. The FCC announced in its 2020 RIF order, “Public safety communications fall into two broad categories: (1) communications within and between public safety entities, and (2) communications between public safety entities and the public.”¹⁵⁸ This narrow definition excludes from the conception of public

155. *ACA Connects v. Bonta*, 24 F.4th 1233, 1237 (9th Cir. 2022).

156. *2020 RIF Order*, *supra* note 12, at 12336, ¶¶ 18–20.

157. *Id.* at 12336, ¶ 20.

158. *Id.* at 12338, ¶ 23.

safety videos and internet uses not mediated by traditional public safety agencies.

The Broadband Institute of California at Santa Clara University School of Law (BBIC) objected to the FCC's vague and apparently narrow definition of public safety in its *Mozilla v. FTC* remand notice.¹⁵⁹ The FCC's statutory duty to promote public safety is not confined to serving institutional public safety agencies. The FCC's statutory mandate is "promoting safety of life and property through the use of wire and radio communications."¹⁶⁰

The FCC's institutional-centric view of public safety would exclude Darnela Fraser's video of Minneapolis Police Officer Derek Chauvin's murder of George Floyd,¹⁶¹ videos of Marjorie Stoneman Douglas High School students during the shootings in Parkland,¹⁶² videos by the victims of the 2018 Camp Fire sparked by PG&E,¹⁶³ and similar internet-based communications vital to public safety.¹⁶⁴ The FCC's definition also excludes communications by critical infrastructure such as electric, natural gas, or water utilities about imminent blackouts, the need to save power or water, or safety tips, even when a life or death is at hand. Neither does it include internet-enabled communications using home Wi-Fi by connected thermostats and

159. *Id.* at 12338. ¶ 23.

160. Meredith Deliso, *Darnella Frazier, who recorded video of George Floyd's death, recognized by Pulitzer board*, ABC NEWS (June 11, 2021, 11:24 AM), <https://abcnews.go.com/us/darnella-frazier-recognized-pulitzer-prizes-george-floyd-video/story?id=78225202>; *New Video Appears to Show George Floyd Being Kneeled On By 3 Officers*, CNN (May 29, 2020), <https://www.cnn.com/videos/us/2020/05/29/george-floyd-kneeled-on-by-three-officers-video-vpx.cnn>.

161. See Abby Ohlheiser & Kayla Epstein, *Just Try to Keep Calm, How One Parkland Student's Phone became his Lifeline and his Voice*, WASH. POST (Mar. 3, 2018), https://www.washingtonpost.com/graphics/2018/lifestyle/parkland-shooting-in-social-media/?utm_term=.07ddba89af90; see also Brandon Griggs, *Hiding Under a Desk as a Gunman Roamed the Halls, a Terrified Student Live-Tweeted a School Shooting*, CNN (Feb. 15, 2018), <https://www.cnn.com/2018/02/15/us/student-live-tweeting-floridaschool-shooting-trnd/index.html>; CNN, *supra* note 161.

162. *Former Firefighter Films [as] He Evacuates Burning Paradise During Camp Fire*, ABC NEWS (Dec. 7, 2018), <https://abc7news.com/video-former-firefighter-films-he-evacuates-burning-paradiseduring-camp-fire/4853479/>; see MICHAEL RAMSEY, THE CAMP FIRE PUBLIC REPORT: A SUMMARY OF THE CAMP FIRE INVESTIGATION 4 (June 16, 2020), <https://www.buttecounty.net/portals/30/cfreport/pge-the-camp-fire-public-report.pdf?ver=2020-06-15-190515-977> (reporting that PG&E plead guilty as charged to all 85 counts of Butte County's criminal indictment alleging 84 individual felony counts of involuntary manslaughter and one count of unlawfully and recklessly causing the Camp Fire as a result of its gross negligence in maintaining its power line).

163. Sandoval, *Net Neutrality Repeal Rips Holes in the Public Safety Net*, *supra* note 9, at 1001.

distributed energy resources used to facilitate energy demand response to stave off blackouts and forestall climate change.¹⁶⁵

The FCC brushed aside as speculative public safety concerns about ISP throttling and net neutrality repeal raised in comments of the Broadband Institute of California at Santa Clara University School of Law (BBIC), the Greenlining Institute, the County of Santa Clara, and Public Knowledge.¹⁶⁶ The FCC's 2018 and 2020 repeal of net neutrality rules adopted in 2015 declined to recognize ISP throttling as a net neutrality violation or threat to public safety.¹⁶⁷

The FCC observed that as of the RIF Remand Order's adoption in October 2020, "all major ISPs have made written commitments not to engage in practices considered to violate open [i]nternet principles, including blocking and throttling."¹⁶⁸ The FCC characterized the terms of those commitments as enforceable by the FTC pursuant to the Memorandum of Understanding (MOU) signed between the Commission and the FTC.¹⁶⁹ That MOU provides the FTC will

investigate and take enforcement action as appropriate against [i]nternet service providers for unfair, deceptive, or otherwise unlawful acts or practices, including . . . actions pertaining to the accuracy of the disclosures such providers make pursuant to

164. Sandoval, *Net Neutrality Powers Energy and Forestalls Climate Change*, *supra* note 9, at 18 ("IoT proliferation illustrates the distributed energy ecosystem."); BBIC, *Comments on Mozilla Remand*, *supra* note 5, at 7–8.

165. *2020 RIF Order*, *supra* note 12, at 12328, ¶ 51 n. 207 (string citing comments of the BBIC, the Greenlining Institute, and Public Knowledge); *infra* note 177.

166. BBIC, *Comments on Mozilla Remand*, *supra* note 5, at 5 ("[T]he FCC's Remand Notice unlawfully attempts to narrow the FCC's responsibilities by considering only an undefined category of "public-safety communications during emergencies."). I served as the lead author for BBIC's comments prepared in collaboration with my colleague Professor Allen S. Hammond, IV and BBIC interns and SCU Law students: Kasey Kagawa, Robert Murillo, Rosa Rico, Ben Katzenberg, and Yi Lu.

167. 47 U.S.C. § 151.

168. *2020 RIF Order*, *supra* note 12, at 12356, ¶ 50 ("We disagree with commenters who assert that the *Restoring Internet Freedom Order* will lead to ISPs engaging in blocking and throttling practices that harm public safety.") The commenters the FCC referenced included the BBIC. *Id.* (citing BBIC, *Comments on Mozilla Remand*, *supra* note 5, at 14 (expressing concern that "current ISP practices including throttling those who use more than certain quantities of data to 2G speeds will interfere with education, public health access, and undercut public safety").

169. *Id.*

170. *Id.*

the Internet Freedom Order’s requirements, as well as their marketing, advertising, and promotional activities.¹⁷⁰

Despite ISP commitments not to engage in throttling, many have contract terms purportedly allowing them to do just that. As BBIC’s comments highlighted, several major wireless ISPs reserved the right to throttle consumers to 2G or other undefined levels when consumers used certain amounts of data.¹⁷¹ Other ISPs reserve the contractual right to throttle consumers who use “excessive” data, at an undefined level.¹⁷²

This Article urges the FCC to examine whether these terms are sufficiently disclosed, consistent with FCC disclosure policy. It also recommends that the FTC review whether such terms violate the FTCA’s proscriptions against deceptive conduct.

On remand from *Mozilla v. FCC*, the FCC’s 2020 OIO relied in large part on ISP disclosures and market information to police internet openness. The FCC determined that even in the absence of ISP commitments to forswear from blocking and throttling, “it is likely that ‘any attempt by ISPs to undermine the openness of the [i]nternet would be resisted by consumers and edge providers.’”¹⁷³ In reaching this conclusion, the FCC failed to analyze the sufficiency of ISP disclosures about their throttling practices or to recognize them as net neutrality or transparency issues.

The FCC concluded, “[c]onsequently, ISPs lack an economic incentive to engage in practices such as blocking or throttling, especially when these practices may harm public safety.”¹⁷⁴ Lack of disclosure about the extent and duration of ISP throttling and its applications to individuals, families, and neighborhoods mutes the ability of public opprobrium to constrain ISP behavior.

The D.C. Circuit noted the limits of public opinion and post-hoc action to prevent or address public safety harms.

Any blocking or throttling of these [i]nternet communications during a public safety crisis could have dire, irreversible results. “[E]ven if discriminatory practices might later be addressed on a

171. *Id.* ¶ 39 (citing Restoring Internet Freedom FCC-FTC Memorandum of Understanding at 2 (Dec. 14, 2017), https://www.ftc.gov/system/files/documents/cooperation_agreements/fcc_fcc_mou_internet_freedom_order_1214_final_0.pdf).

172. BBIC, *Comments on Mozilla Remand*, *supra* note 5, at 45–48.

173. 2020 RIF Order, *supra* note 12, at 12328. See Brodtkin, *supra* note 4.

174. 2020 RIF Order, *supra* note 12, at 12328 (citing Restoring Internet Freedom Order, 33 FCC Rcd. at 396, ¶ 142).

post-hoc basis by entities like the Federal Trade Commission,” the harm to the public “cannot be undone.”¹⁷⁵

The FCC disagreed “with commenters who assert that the *Restoring Internet Freedom Order* will lead to ISPs engaging in blocking and throttling practices that harm public safety.”¹⁷⁶ The FCC effectively classified throttling as speculative, ignoring evidence presented of ISP throttling behavior.¹⁷⁷ This conclusion rests on regulatory semiotics¹⁷⁸ that fail to classify ISP throttling based on vague contract terms as the type of throttling net neutrality rules sought to curtail.

Through a process Professor Becky Lentz calls “linguistic engineering,”¹⁷⁹ the FCC defined throttling as neither a public safety problem, net neutrality problem, nor a regulatory problem. The FCC’s conceptualization of throttling as speculative and neither a safety nor net neutrality problem is intertwined with the FCC’s definition of public safety as mediated through public safety institutions. The FCC failed to recognize communications between the citizenry or between the public and institutions designed as Critical

175. *Id.*

176. *Id.*

177. *Id.* at 12356, ¶ 50 n. 201 (citing BBIC, *Comments on Mozilla Remand*, *supra* note 5, at 14 (expressing concern that “current ISP practices including throttling those who use more than certain quantities of data to 2G speeds will interfere with education, public health access, and undercut public safety”)); BBIC, Reply Comments, In the Matter of Restoring Internet Freedom (WC Docket Nos. 17-108, 17-287, 11-42), at 72 (May 20, 2020), <https://ecfsapi.fcc.gov/file/1052104890909/BBIC%20Mozilla%20Remand%20Reply%20Comments%20Final.pdf> (“The FCC’s Title I classification allows ISPs to demand extra payments for priority transmission or to be safeguarded from the ISPs intentional acts that manipulate priority.”); Reply Comments of the Greenlining Institute on February 19, 2020 Public Notice at 7 (May 20, 2020), <https://ecfsapi.fcc.gov/file/10520282663174/Docket%20No.%2017-108%2C%20Reply%20Comments%2C%20The%20Greenlining%20Institute.pdf> (“‘Public safety communications’ include an extremely broad array of activities and services, and providers’ narrow interpretation could lead to the blocking or throttling of online activity that is critical to protecting public safety. The consequences of this blocking or throttling could result in injury or death.”); Comments Of Public Knowledge, Access Humboldt, Access Now, And National Hispanic Media Coalition, at 8, https://ecfsapi.fcc.gov/file/10420030611321/PK_et_al_net_neutrality_remand_4-20-2020.pdf (“The FCC’s reclassification leaves the FCC with no corrective tool to prevent carriers from throttling emergency services.”).

178. See Barton Beebe, *The Semiotic Analysis of Trademark Law*, 51 UCLA L. REV. 621, 631 (2004) (“Semiotics investigates the relation between ‘structure’ (a language’s underlying system of rules and regularities) and ‘event’ (specific language uses). . . . Event assumes and informs structure and structure assumes and informs event.”).

179. Roberta Lentz, *Regulation as Linguistic Engineering*, in THE HANDBOOK OF GLOBAL MEDIA AND COMMUNICATION POLICY 432, 435, 439 (Robin Mansell & Marc Raboy eds., 2011).

Infrastructure pursuant to CIPA (such as schools and healthcare providers) as public safety communications. As the Santa Clara County Fire Department incident exemplifies, the FCC's failure leaves all of us at risk.

The FCC's government-mediated public safety framework created the basis for linguistic engineering which side-stepped analysis of ISP throttling practices on public safety and the diverse public, including communities of color. The FCC failed to recognize that ISP policies which slow users to 2G speeds disable meaningful access to many health and critical infrastructure services, such as education, by making many common applications and platforms inaccessible. ISP throttling limits a "user's ability to select, access, and use broadband [i]nternet access service or the lawful [i]nternet content, applications, services, or devices of the end user's choice," without targeting traffic based on internet content, application, or service.¹⁸⁰

The FCC's 2020 RIF Order did not argue or determine that throttling users to 2G speeds after using internet bandwidth necessary for several days of digital work or school was "reasonable network management" under the 2015 OIO's rules.¹⁸¹ The reasonableness of throttling practices remains largely unexamined.

FCC transparency rules still standing after the FCC's 2018 net neutrality repeal and 2020 RIF Order create a means to examine whether ISP disclosures about practices that slow users to 2G speeds sufficiently inform the public about the nature of internet access offered. Transparency rules assume increased urgency as the FCC's EBB program rules rely on consumer choice to achieve the program's objectives.¹⁸²

D. THE EMERGENCY BROADBAND BENEFIT PROGRAM

Congress established the EBB program on December 27, 2020, through the Consolidated Appropriations Act, 2021, to provide relief during the COVID-19 pandemic. Through the EBB, "eligible low-income households may receive a discount off the cost of broadband service and certain connected devices during an emergency period relating to the COVID-19

180. *Cf.* California S.B. 822, Section (7)(A).

181. 2020 RIF Order, *supra* note 12, at 12357, ¶ 52 ("[W]e find unpersuasive commenters' concerns regarding the effect of service plans that limit data or speeds on members of the public who rely on mass market broadband Internet access services to access public safety information. We observe that broadband service plans that limit data or speeds were not prohibited even under the *Title II Order*; as such, we find the return of broadband Internet access service to its information services classification and elimination of the conduct rules irrelevant to the impact on the permissibility of throttling under a data plan when the data cap is exceeded."); *cf.* FCC, 2015 Open Internet Order, *supra* note 58, at 5604.

182. FCC, EBB Order, *supra* note 70, ¶ 73.

pandemic, and participating providers can receive a reimbursement for such discounts.”¹⁸³ Participating EBB providers “will make available to eligible households a monthly discount off the standard rate for an [i]nternet service offering and associated equipment, up to \$50.00 per month” and up to \$75.00 per month on Tribal Lands.¹⁸⁴ The bill delegated responsibility to the FCC to develop the EBB program rules and standards.

The FCC did not adopt EBB minimum service standards. Its order determined that qualifying internet service offerings must permit “households to rely on these connections for the purposes essential to participating in society during the pandemic, such as telework, remote learning, and telehealth.”¹⁸⁵ An EBB qualifying internet service offering is defined as “broadband internet access service provided by such provider to a household, offered in the same manner, and on the same terms, as described in any of such provider’s offerings for broadband internet access service to such household, as of December 1, 2020.”¹⁸⁶ Effectively, under the EBB, authorized providers could receive federal support for plans offered as of December 1, 2020.

The FCC requires ISPs to “disclose accurate information regarding the performance characteristics, commercial terms, and other features of their discounted broadband services to enable consumers to make informed choices regarding the purchase and use of such services.”¹⁸⁷ Lack of accurate information, inconspicuous information, or terms that are functionally unintelligible undermine disclosure, consumer choice, and competition. Failure to explain 2G speed characteristics and identify the types of programs, applications, and platforms inaccessible at that throttled speed¹⁸⁸ exemplify inadequate disclosure that undermines ROBIN access.

183. *Id.* ¶ 2.

184. *Id.* ¶ 4.

185. *Id.* ¶ 73.

186. *Id.* ¶ 72.

187. *Id.* ¶ 73 (citing 47 C.F.R. § 8.1(a) (the FCC ISP transparency rule adopted in 2018)); see also Douglas A. Hass, *The Never-Was-Neutral Net and Why Informed End Users can End the Net Neutrality Debates*, 22 BERKELEY TECH. L.J. 1565, 1630 (2007) (proposing ISP service offering and traffic control policy disclosure as an alternative to net neutrality regulations); Christopher S. Yoo, *What can Antitrust Contribute to the Network Neutrality Debate?*, 1 INT’L J. COMM. 493, 529 (2007) (advocating better ISP disclosure of policies and internet application limits in lieu of regulation or net neutrality legislation).

188. See *infra* notes 227–32 and accompanying text.

IV. FTC ACT DECEPTIVE CONDUCT PROSCRIPTIONS

The FCC's 2018 decision to reclassify ISPs under Title I gave ISP jurisdiction back to the FTC.¹⁸⁹ The FTC Act proscribes deceptive conduct under its statutory mandate to prevent “unfair methods of competition” and “unfair or deceptive acts or practices in or affecting commerce”¹⁹⁰ Section 5(a) of the FTC Act provides that “unfair or deceptive acts or practices in or affecting commerce . . . are . . . declared unlawful.”¹⁹¹

The FTC's 1984 Policy Statement on Deception¹⁹² defines deceptive practices “as involving a material representation, omission or practice that is likely to mislead a consumer acting reasonably in the circumstances.”¹⁹³ An act or practice is “unfair” if it “causes or is likely to cause substantial injury to consumers which is not reasonably avoidable by consumers themselves and not outweighed by countervailing benefits to consumers or to competition.”¹⁹⁴ “An act has been held to be deceptive if it involves a material representation, omission, or practice that is likely to mislead consumers acting reasonably under the circumstances.”¹⁹⁵

Deception claims often focus on whether advertisements omit material information or are misleading. The FTC Policy Statement on Deception lists examples of practices that have been found to be misleading or deceptive, including “false oral or written representations” and “failure to perform promised service.”¹⁹⁶ Former FTC Commissioner Joshua D. Wright and his co-author Jay S. Kaplan observed that under the section 5 of the FTCA, there “is no need to balance any potential benefits, because, under the deception

189. FCC, *2018 RIF Order*, *supra* note 12, at 7852, 7878 (classifying ISPs under Title I returns jurisdiction over ISPs to the FTC, “the nation’s premier consumer protection agency,” to police ISP privacy, anticompetitive acts, or unfair and deceptive practices).

190. 15 U.S.C. §§ 41–77.

191. 15 U.S.C. § 45(a)(1).

192. FED. TRADE COMM’N, FTC POLICY STATEMENT ON DECEPTION (Oct. 14, 1983), https://www.ftc.gov/system/files/documents/public_statements/410531/831014deceptionstmt.pdf (appended to *Cliffdale Assocs., Inc.*, 103 F.T.C. 110, 174 (1984)).

193. *A Brief Overview of the Federal Trade Commission’s Investigative, Law Enforcement, and Rulemaking Authority*, (Oct. 2019), <https://www.ftc.gov/about-ftc/what-we-do/enforcement-authority>.

194. *Id.* (citing 15 U.S.C. § 45(n)).

195. Sandoval, *Disclosure, Deception and Deep-Packet Inspection*, *supra* note 4, at 662 (citing *FTC v. Cyberspace.com LLC*, 453 F.3d 1196, 1199 (9th Cir. 2006); *accord* *FTC v. Tashman*, 318 F.3d 1273, 1277 (11th Cir. 2003); *Telebrands Corp.*, 140 F.T.C. 278, 290 (2005), *aff’d*, 457 F.3d 354 (4th Cir. 2006); *Cliffdale Assocs.*, 103 F.T.C. at 164–65).

196. FTC POLICY STATEMENT ON DECEPTION, *supra* note 192.

prong, there are no cognizable benefits when an ISP blatantly misleads consumers.”¹⁹⁷

In 2014 the FTC filed a lawsuit alleging that AT&T violated the FTC Act’s deceptive conduct provisions through its “unlimited” data plan advertisements that were inconsistent with its practice of throttling customers who used between 2–3 gigabits (GB) of data in a billing cycle (a month-long period) to speeds as low as 128 kilobits per second (kbps).¹⁹⁸ AT&T filed a motion to dismiss the FTC’s claim, arguing that since its mobile phone service provided both voice telephone services as a common carrier and internet services, its status as a common carrier for voice service precluded FTC jurisdiction over its internet services provided through that same mobile phone.¹⁹⁹

Congress created a common-carrier exception to the FTC’s enforcement power in deference to the Communications Act of 1934, as amended.²⁰⁰ In *AT&T Mobility v. FTC*, heard *en banc* by the Ninth Circuit, at issue was whether the FTCA “common-carrier exemption is activity-based, meaning that a common carrier is exempt from FTC jurisdiction only with respect to its common-carrier activities, or status-based, such that an entity engaged in common-carrier activities is entirely exempt from FTC jurisdiction.”²⁰¹

The Ninth Circuit affirmed *en banc* the district court’s denial of AT&T’s motion to dismiss. Examining “the FTC Act’s text, the meaning of ‘common carrier’ according to the courts around the time the statute was passed in 1914, decades of judicial interpretation, the expertise of the FTC and [FCC], and legislative history, [the Ninth Circuit] conclude[d] that the exemption is activity-based.”²⁰²

The Ninth Circuit determined that the common-carrier exemption “provides immunity from FTC regulation only to the extent that a common carrier is engaging in common-carrier services.”²⁰³ “A phone company is no

197. Joshua D. Wright & Jay S. Kaplan, *All of That in One Page: The Application of the 2015 FTC Unfair Methods of Competition Policy Statement to Net Neutrality Disputes*, 17 COLO. TECH. L.J. 311, 333 (2019).

198. Complaint for Permanent Injunction and Other Equitable Relief ¶¶ 16–19, *FTC v. AT&T Mobility LLC*, No. 14-CV-04785-EMC (N.D. Cal. 2015), <https://www.ftc.gov/system/files/documents/cases/141028attcmpt.pdf>.

199. *FTC v. AT&T Mobility LLC*, 883 F.3d 848, 850 (9th Cir. 2018).

200. *Id.* at 850 (citing 15 U.S.C. § 45(a)(1), (2)); *FTC v. Verity Int’l, Ltd.*, 443 F.3d 48, 56 (2d Cir. 2006).

201. *AT&T Mobility*, 883 F.3d at 850.

202. *Id.*

203. *Id.*

longer just a phone company,” the Ninth Circuit recognized.²⁰⁴ “The transformation of information services and the ubiquity of digital technology mean that telecommunications operators have expanded into website operation, video distribution, news and entertainment production, interactive entertainment services and devices, home security and more.”²⁰⁵

Millions of Americans depend on “phones” that access the internet and the public switched telephone network to make phone calls.²⁰⁶ “Reaffirming FTC jurisdiction over activities that fall outside of common-carrier services avoids regulatory gaps and provides consistency and predictability in regulatory enforcement.”²⁰⁷

On remand from *FTC v. AT&T Mobility*, the FTC and AT&T entered into a stipulated permanent injunction and monetary judgement that prohibits AT&T from “[m]aking any representation about the amount or speed of mobile data, including that the mobile data is unlimited, without disclosing, Clearly and Conspicuously and in Close Proximity to the representation, all Material Restrictions imposed by Defendant.”²⁰⁸ The monetary judgment collected \$60,000,000 as an escrow account to support refunds to AT&T customers.²⁰⁹ Roslyn Layton and Tom Struble point to the FTC’s AT&T Mobility case as evidence that such “contracts are enforceable under the FTC’s consumer protection authority without any need for ex ante regulation.”²¹⁰

The Supreme Court in *AMG Capital Management, LLC v. FTC* determined that section 13(b) of the FTC Act does not authorize the FTC to “seek, and a court to award, equitable monetary relief such as restitution or disgorgement,” effectively curtailing FTC remedy options.²¹¹ The Court recognizes that the FTCA authorizes the FTC to obtain a “‘permanent injunction’ in federal court against ‘any person, partnership, or corporation’ that it believes ‘is violating, or is about to violate, any provision of law’ that the Commission enforces.”²¹²

204. *Id.* at 851.

205. *Id.*

206. *See* 47 C.F.R. § 54.101(a)(1) (defining supported voice telephony services as providing voice grade access to the public switched network or its functional equivalent).

207. *AT&T Mobility*, 883 F.3d at 851.

208. *FTC AT&T Mobility Stipulated Order*, *supra* note 11, at 4.

209. *Id.* at 5–6.

210. Roslyn Layton & Tom Struble, *Net Neutrality Without the FCC?: Why the FTC Can Regulate Broadband Effectively*, 18 FEDERALIST SOC’ REV. 132, 135 (2017).

211. *AMG Capital Management, LLC v. FTC*, 141 S. Ct. 1341, 1344 (2021) (citing 87 Stat. 592, 15 U.S.C. § 53(b)).

212. *Id.*

The Supreme Court’s textualist approach in *AMG Capital Management* focused on the “text and structure of the particular statutory scheme at issue.”²¹³ That analysis allows the FTC to seek an injunction for FTCA violations but precludes the FTC from seeking monetary relief in federal court for FTCA violations. *AMG Capital Management* leaves in place the FTC’s injunction against AT&T’s representations about the “amount or speed of mobile data,” without clear, conspicuous, and closely proximate disclosure about material restrictions imposed by AT&T, including those which throttle speed.²¹⁴ *AMG Capital Management* allows the FTC to analyze ISP throttling policies and seek an injunction if it believes section 5 of the FTCA is violated but not to seek monetary relief in federal court, including restitution of customer payments or disgorgement of ISP profits.

V. THROTTLING INTERNET ACCESS LIMITS EQUITY AND RISKS PUBLIC HEALTH AND SAFETY

A. INTERNET ACCESS NEEDS ZOOM PAST THE FCC’S ADVANCED SERVICES DEFINITION

This Article recommends the FCC and FTC examine ISP policies that throttle users to 2G speeds, rendering inaccessible many communications platforms vital for education, health, work, economic opportunity, and civic participation. Some ISPs slow users to 1990s-level 2G speeds after internet users consume data commensurate with two weeks of digital law school or one week of high school coursework via videoconferencing.²¹⁵ ISP throttling may leave users unable to access news sources, telemedicine, or videoconferencing used for school, work, and news interviews.²¹⁶

Videoconferencing platforms are often used for telemedicine and require substantial upload and download bandwidth.²¹⁷ Videoconferencing is

213. *Id.* at 1343; Robert J. Pushaw, Jr., *Comparing Literary and Biblical Hermeneutics to Constitutional and Statutory Interpretation*, 47 PEPPERDINE L. REV. 463, 491 (2020) (observing that as of February 2020, five Justices—Roberts, Thomas, Alito, Gorsuch, and Kavanaugh—have formally declared that they are constitutional originalists and statutory textualists).

214. *FTC AT&T Mobility Stipulated Order*, *supra* note 11, at 4.

215. BBIC, *Comments on Mozilla Remand*, *supra* note 5, at 49 (noting that a high school student using “their mobile phone to access online education would hit a 50 GB threshold in a little over a week of full-time classes,” while a law school student would likely reach that cap that triggers ISP slowdowns within two weeks of coursework).

216. *Id.* at 18–19.

217. *Id.* at 18 (citing Teresa Iafolla, *What are the Basic Technical Requirements for Telehealth?* EVISIT, <https://blog.evisit.com/what-are-the-basic-technical-requirements-for-telehealth> (last visited Apr. 20, 2020) (recommending 15 mbps down and 5 mbps up for telehealth visits that include video chat); *Technology Requirements*, AM. ACAD. ALLERGY, ASTHMA, &

increasingly used across many sectors of society, creating a mismatch between the FCC's 2020 definition of "advanced services," ISP throttling to 2G speeds, and contemporary communications needs.

In 2020, the FCC classified internet networks providing 25 megabits per second (mbps) down and 3 mbps up (25/3) speed as "advanced service."²¹⁸ Internet-based education, work, and healthcare access sputter at 25/3 "advanced speeds" as defined by the FCC, leaving many Americans unable to meet modern communications needs.²¹⁹

The bandwidth capacity (measured in GB per hour) needed for videoconferencing through Zoom for activities such as synchronous classes illustrates the risks of throttling users to 2G speeds. A speed test conducted in April 2020 by SCU Law student and BBIC Intern Kasey Kagawa using an Android Galaxy Note 9 cellphone on Zoom found:

A 30-minute Zoom video conference uses 0.35 GB of mobile data use per half-hour, rounded up to 0.4 GB per half-hour = 0.8 GB per hour.

Every 10 hours of class conducted through video conferencing using an Android Galaxy Note 9 consumes 8 Gbs [sic] of data.

When using the Android Galaxy Note 9 as a hot spot, a 30-minute Zoom video conference uses 0.56 GB of mobile data use per half-hour rounded up to 0.6 GB per half-hour = 1.2 GB per hour.

Every 10 hours of class on video conferencing using an Android phone as a Mobile hotpot uses 12 Gbs [sic] of data.

15 hours a week of class using Zoom video conferencing through an Android phone as a Mobile hot spot will use approximately 18 Gb [sic] of data.

With other research projects and necessary meetings with students or professors to prepare for presentations, many law students will

IMMUNOLOGY, <https://www.aaaai.org/Allergist-Resources/Telemedicine/technology> (last visited April 20, 2020 ("Most basic to a telemedicine practice is a secure broadband internet connection. The amount and speed of the internet connection will determine the video quality and amount and speed of data transfer. A basic business broadband connection should be sufficient at about 50–100 Mbps (megabits per sec)"))

218. BROADBAND DEPLOYMENT REPORT, *supra* note 91, at 8.

219. See Jon Brodtkin, *FCC chair proposes new US broadband standard of 100Mbps down, 20Mbps up*, ARS TECHNICA, July 15, 2022, <https://arstechnica.com/tech-policy/2022/07/fcc-chair-proposes-new-us-broadband-standard-of-100mbps-down-20mbps-up/> ("The needs of Internet users long ago surpassed the FCC's 25/3 speed metric, especially during a global health pandemic that moved so much of life online," FCC Chairwoman Rosenworcel said in announcing a proposal to increase the FCC's advanced Internet speed definition).

use approximately 50 GB of data every two weeks if using an Android phone as a Mobile hot spot to access online education.²²⁰

Many high school classes meet 5–6 hours a day.²²¹ Students using their mobile phone to access online education for high school classes conducted full-time would hit a 50 GB threshold in a little over a week of full-time classes.²²²

Zoom explains that it requires 3G or 4G speeds.²²³ For group video calling such as the sessions used for online education and some telemedicine applications, Zoom requires speeds of 1.0 mbps up and 600 kbps down, and 2–4 times more speed to receive gallery views, depending on the number of participants.²²⁴ To receive “1080p HD video” Zoom requires 3.0 Mbps down and 3.8 Mbps up.²²⁵ Many webcams use the 1080p HD video standard for smooth video transmission, though 4K video offers a higher quality standard but requires more upload and download bandwidth.²²⁶

With the expansion of faster networks including LTE and 5G networks, “2G speeds are actually imposed by software on the network. These 2G speeds range from 256Kbps, to 128Kbps, down to 64Kbps.”²²⁷ Stetson Doggett explains:

[256 kbps] is probably the slowest speed you can get by on if you’re just doing some light email, text-based chatting like Facebook Messenger, WhatsApp, or iMessage, or streaming music in the background. Expect things to take longer, but if you’re patient enough they will load.

At 128Kbps and 64Kbps, we have things start to basically break down. Most notably, music streaming. Instead of getting a smooth playback, music streaming was choppy. It was impossible to listen

220. BBIC, *Comments on Mozilla Remand*, *supra* note 5, at 49.

221. *Id.* (citing Table 5.14. *Number of instructional days and hours in the school year, by state: 2018*, NAT’L CTR. EDUC. STATISTICS, https://nces.ed.gov/programs/statereform/tab5_14.asp).

222. *Id.*

223. ZOOM, *Zoom system requirements: Windows, macOS, Linux* (last updated May 23, 2022), https://support.zoom.us/hc/en-us/articles/201362023-Zoom-system-requirements-Windows-macOS-Linux#h_d278c327-e03d-4896-b19a-96a8f3c0c69c (“System requirements, An internet connection—broadband wired or wireless (3G or 4G/LTE)”).

224. *Id.* (“For gallery view receiving: 2.0Mbps (25 views), 4.0Mbps (49 views).”).

225. *Id.*

226. Luke Edwards, *Best Webcams for Teachers and Students 2020*, TECH LEARNING <https://www.techlearning.com/buying-guides/best-webcams-for-teachers-and-students-2020> (last visited May 23, 2021) (recommending “the best webcams for teachers and students during remote learning” and advising that “to ensure you have the best scalable image a 1080p camera could be worth the investment”).

227. Stetson Doggett, *How Fast are Capped 2G Speeds? LTE vs 3G vs 2G Data Speed Test*, BEST PHONE PLANS (Mar. 14, 2021), <https://www.bestphoneplans.net/news/2g-speed-test>.

to even a single song without the music pausing to load and catch up with itself.²²⁸

Zoom requires 3G or 4G speeds for synchronous video such as telemedicine or classes via videoconferencing.²²⁹

ISPs that throttle users to 2G speeds of 128 kbps leave subscribers without sufficient speed to maintain access to Zoom videoconferencing. ISP throttling of users to 2G speeds, the standard launched in 1991, will likely cause videoconferencing and synchronous or asynchronous video uses to fail altogether, not just be choppy.²³⁰ Throttling internet users in this fashion renders health video visits and many platforms used for digital school and work inaccessible, raising public health and safety risks. At 2G speeds, there is no zooming internet access, only a slow-motion crawl. Zoom and many common internet applications used for contemporary school, work, and health uses will not operate at 2G speeds.

Connecticut Attorney General William Tong reported that his office heard from Connecticut families who easily exceeded Comcast's data cap in 2020 because they relied on videoconferencing to attend school and work remotely. "Far from so-called super users, these were stories from typical Connecticut families merely trying to stay employed and educate their children during a global pandemic."²³¹ Although data caps are distinct from throttling, a common practice is for ISPs to throttle their users' internet speed after users hit certain data use levels or data caps.²³²

The Student Home Internet Connectivity Study examined internet connection information about 750,000 students from urban, suburban, and rural districts in thirteen school districts that provided network logs and other information about student connections to schoolwork.²³³ That study found

228. *Id.*

229. ZOOM, *supra* note 219.

230. BBIC, *Comments on Mozilla Remand*, *supra* note 5, at 19 (citing Ashutosh Bhatt, *Difference Between 2G and 3G Technology*, ENGINEERS GARAGE, (Mar. 27, 2012), https://www.engineersgarage.com/how_to/difference-between-2g-and-3g-technology/).

231. Jon Brodtkin, *Comcast Reluctantly Drops Data-cap Enforcement in 12 States for Rest of 2021*, ARS TECHNICA (Feb. 19, 2021), <https://arstechnica.com/tech-policy/2021/02/comcast-responds-to-pressure-cancels-data-cap-in-northeast-us-until-2022/>.

232. *Id.*; *Is My Internet Being Throttled by My ISP?*, INTERNET SERV. PROVIDERS (Mar. 26, 2020), <https://www.isp.com/blog/is-my-internet-being-throttled-by-my-isp/> ("Data caps are one of the most common reasons for ISP throttling. A data cap is the maximum amount of data you are allowed to download or stream in a single month. While some major providers, such as Spectrum and Verizon Fios, don't have data caps, many others do. For example, Comcast throttling may occur after you hit 1 TB of data usage in a month.").

233. COSN, *supra* note 28, at 6 ("Each participating school district provided data such as student characteristics, network logs, Quality of Service (QoS) data for meeting software,

the FCC's definition of "advanced services" as 25/3 "is inadequate to support even a single student in a household, let alone multiple students."²³⁴ "[O]ver 70% of students live in a household with one or more other students. Concurrently supporting multiple students using video from the same internet connection is problematic when bandwidth availability is low."²³⁵

"Home network bandwidth capacity must account for concurrent usage by multiple students, including current video use."²³⁶ CoSN (the Consortium for School Networking) recommends "a per-student minimum bandwidth standard of a download speed of 25 Mbps and upload speed of 12 Mbps to support concurrent activity and usage."²³⁷ "Given the new requirements of videoconferencing for classroom communication and student collaboration, ISPs receiving federal support should provide unlimited data for home learning connections without throttling," the Student Home Internet Connectivity Study recommended.²³⁸ The pandemic has changed the internet use paradigm, requiring law and regulation to catch up.

The Student Home Internet Connectivity Study found that "regardless of the student's [internet protocol (IP)] address, 92% of students in the study connected to the internet via WiFi instead of a wired connection."²³⁹ It also found that "many students shared an IP address with other students that were not from the same household. Likely causes include students wanting social interaction with other kids, finding a faster internet connection at a friend's house, and parents who share childcare responsibilities."²⁴⁰

In 2018, approximately 79.7 million Americans used prepaid mobile internet plans, while 273.9 million subscribed to postpaid plans.²⁴¹ Prepaid mobile use was spread nearly evenly by age in 2018 with 29.5% of those age 18–29, 32.7% of those age 30–49, and 25.5% of those age 50–64 using prepaid plans.²⁴² Pew reports that "[r]eliance on smartphones for online

Internet Service Provider (ISP) data, and geolocation data. Thirteen urban, suburban, and rural school districts representing approximately 750,000 students from across the United States participated in the study over the course of six weeks.").

234. *Id.* at 8.

235. *Id.* at 10.

236. *Id.*

237. *Id.* at 8.

238. *Id.* at 11.

239. *Id.* at 12.

240. *Id.*

241. Jason Leigh, *U.S. Postpaid and Prepaid Wireless Forecast, 2019–2023*, IDC (Dec. 2019), <https://www.idc.com/getdoc.jsp?containerId=US44687219>.

242. *Share Of Americans Using Prepaid Service for Their Cell Phone in 2018, By Age*, STATISTA (2021), <https://www.statista.com/statistics/231638/cell-phone-users-who-use-a-prepaid-card-usa/>.

access is especially common among younger adults, lower-income Americans and those with a high school education or less.”²⁴³ Smartphone-only internet access is also more common for Latinx and African American people than for Whites, Pew reported.²⁴⁴

Not all students, whether in K-12 or higher education, have a wired internet connection or the ability to order or pay for wired internet. Nationwide in 2020, “across all racial and ethnic groups, 16.9 million children remain logged out from instruction because their families lack[ed] the home internet access necessary to support online learning.”²⁴⁵ The Alliance for Excellent Education found that for Black, Latinx, and American Indian/Alaska Native households, one out of three lacked home internet access.²⁴⁶

The pandemic underscored the fact that millions of Americans are relegated to the status of “internet migrants,” able to access the internet only at library steps or a fast-food parking lot.²⁴⁷ Communications regulation needs to serve diverse American communities, not just the imagined community of homeowners, high wage earners, and those able to install wired internet.

Many people access the internet through mobile phones or hot spots for vital services such as telemedicine, counseling, education, social and economic services, work, and civic organization. Even for those who have wired internet access, events such as outages, intermittency, fire evacuation, flood, other hazards, or housing insecurity have forced many to switch to mobile phones as a hotspot to support computer access during online classes.²⁴⁸

The COVID-19 pandemic increased housing insecurity, diminishing access to wireline internet and deterring consumer investment in such services. “According to the Census Bureau’s Household Pulse Survey for late

243. *Mobile Fact Sheet*, PEW RSCH. CTR. (Apr. 7, 2021), <https://www.pewresearch.org/internet/fact-sheet/mobile/?menuItem=d40cde3f-c455-4f0e-9be0-0aefcdacee00>.

244. *Id.*

245. ALLIANCE EXCELLENT EDUC., STUDENTS OF COLOR CAUGHT IN THE HOMEWORK GAP 1 (2020), https://futureready.org/wp-content/uploads/2020/08/HomeworkGap_FINAL8.06.2020.pdf.

246. *Id.* at 2 (“Thirty-four percent of American Indian/Alaska Native families and about 31 percent each of Black and Latino families lack access to high-speed home internet compared to only about 21 percent of White families.”).

247. Catherine J.K. Sandoval, *Net Neutrality Protects Public Safety*, Presentation at Santa Clara University High-Tech Law Institute Symposium, Promoting Safety of Life and Property Through the Open Internet, (Mar. 25, 2020), <https://1x937u16qcra1vnejt2hj4jl-wpengine.netdna-ssl.com/wp-content/uploads/Professor-Sandoval-Net-Neutrality-Protects-Public-Safety-SCU-Law-HTLI-Webinar-March-25-2020.pdf>; Rachel Fried, *Compounding Crises Technology and Equity in the COVID Era*, 249 J. COLL. ADMISSION 15, 16 (2020).

248. See BBIC, *Comments on Mozilla Remand*, *supra* note 5, at 48.

September [2020], renters earning less than \$25,000 a year were much more likely to report lost employment income since the March shutdown.”²⁴⁹ As a result, “one in five renters earning less than \$25,000 also said they were behind on rent, compared with 15 percent of all renters and just 7 percent of renters earning more than \$75,000.”²⁵⁰ Due to COVID-related income losses “23 percent of Black, 20 percent of Hispanic, and 19 percent of Asian renters were behind on their rents by late September, or about twice the 10 percent share of white renters.”²⁵¹

Deutsche Bank Research used mobile phone geolocation data from New York, Los Angeles, and Chicago during the COVID-19 pandemic in 2020 to track neighborhood levels of mobility and those who were able to stay at home.²⁵² Their study found that

Blacks had to venture out of their homes 135% more than Whites compared to pre-Covid levels, during the riskiest of times across the cities. [Deutsche Bank Research] believe[d] this [wa]s an accurate representation of the state of the racial digital divide in the country. Clearly, poor access to Tech connectivity & work-from-home jobs rendered minorities with few choices but to venture out of home to make a living, even with peril to their lives.²⁵³

According to Dr. Mary Bassett, Director of the FXB Center for Health and Human Rights at Harvard University, “a big determinant of who dies is who gets sick in the first place, and that infections have been far more prevalent among people who can’t work from home.”²⁵⁴ Throttling can limit the ability of people to work, study, shop for essentials, access healthcare, and engage in civic activities at home, increasing public health and safety risks.

The COVID-19 pandemic laid bare the gulf between reality and models designed to serve the imagined community of internet users able to install robust internet access in a home whose physical access they control.²⁵⁵ That regulatory imagination did not adequately envision the needs of Americans

249. JOINT CTR. HOUS. STUD. HARVARD UNIV., THE STATE OF THE NATION’S HOUSING 2020 1 (2020), https://www.jchs.harvard.edu/sites/default/files/reports/files/Harvard_JCHS_The_State_of_the_Nations_Housing_2020_Report_Revised_120720.pdf.

250. *Id.*

251. *Id.* at 2.

252. WALIA & RAVINDRAN, *supra* note 37, at 5.

253. *Id.*

254. Sandoval et al., *supra* note 31, at 425.

255. See Sandoval & Lanthier, *supra* note 18, at 16; ADAM BANKS, RACE, RHETORIC, AND TECHNOLOGY: SEARCHING FOR HIGHER GROUND 41 (2005); cf. BENEDICT ANDERSON, IMAGINED COMMUNITIES: REFLECTIONS ON THE ORIGIN AND SPREAD OF NATIONALISM (revised ed. 1991).

who live in households with several children and adults, each needing internet bandwidth for school, work, healthcare, and other vital services. Neither did it imagine the needs of those living in multi-generational or multi-household dwellings, those with insecure housing or no housing, nor those facing eviction, fire, flood, or other dangers.

B. THROTTLING TO THE '90S

In 2015, the FCC issued a \$100 million Notice of Apparent Liability (NAL) against AT&T for violations of the 2010 transparency rules for slowing customers on “unlimited” data plans to speeds where mapping and other common applications would not work.²⁵⁶ In its 2017 brief to the Ninth Circuit, the FCC noted that a “majority of the FCC’s current commissioners dissented from the decision to issue the NAL . . . and no further action has been taken on it.”²⁵⁷

Six years later during the COVID-19 pandemic, despite the FTC permanent injunction ordering AT&T to prominently disclose any limitations on internet use in a way that informs customers about speed limits, AT&T’s throttling disclosures leave consumers ill-informed and raise FTCA compliance issues. This issue is not unique to AT&T. Many ISP websites and representations make it difficult to find or understand the limitations customers will face and the activities their ISP will limit.

The FTC and several states sued Frontier for FTCA deceptive conduct violations, alleging that “Frontier did not provide many consumers with the maximum speeds they were promised and the speeds they actually received often fell far short of what was touted in the plans they purchased.”²⁵⁸ The Frontier case analyzed representations of speeds “up to” stated levels,²⁵⁹ but it was not a throttling complaint. The Frontier and AT&T FTC cases highlight the FTC’s potential to seek and impose injunctions against ISPs for misrepresentation of internet access levels provided. Examination of ISP contract terms in 2021 reveals the urgency of throttling analysis.

256. In the Matter of AT&T Mobility, LLC., 30 FCC Rcd. 6613 (2015).

257. BBIC, *Comments on Mozilla Remand, supra* note 5, at 45 (citing Brief of the Federal Communications Commission as Amicus Curiae in Support of Plaintiff-Appellee at 3, FTC v. AT&T Mobility LLC, 2017 WL 2398744 (No. 15-16585) (9th Cir. 2017)).

258. Press Release, FTC Sues Frontier Communications For Misrepresenting Internet Speeds (May 19, 2021), <https://www.ftc.gov/news-events/press-releases/2021/05/ftc-sues-frontier-communications-misrepresenting-internet-speeds>.

259. Complaint for Preliminary Injunction, Permanent Injunction, Monetary Relief and Other Relief ¶ 44, FTC v. Frontier, No. 2:21-cv-4155 (C.D. Cal. filed May 19, 2021). https://www.ftc.gov/system/files/documents/cases/dkt_1_-_complaint.pdf.

In May 2021, T-Mobile's website stated in smaller print than its prepaid plan description that T-Mobile will throttle users to 128 kbps after they consume 5 GB of data or 3 GB for tethering:

On qualifying Prepaid ONE Plan, in Canada/Mexico, up to 5GB high-speed data, then unlimited at up to 128kbps. Tethering at max 3G. For the small fraction of customers using more than 50GB per month, primary data usage must be on smartphone or tablet. Simply Prepaid & Mobile Internet: Domestic use only; additional charges apply for international use, where available. Partial minutes/megabytes rounded up. Full speeds available up to data allotment, including tethering; then slowed to up to 2G speeds for balance of service period.²⁶⁰

T-Mobile's page advertising prepaid plans did not define 2 G speeds, nor did it contain any links defining those speeds or explain what types of internet uses 128 kbps or 2 G speeds will render inaccessible.

After 6.25 hours of class via Zoom videoconference, users of T-Mobile's prepaid plan would hit the 5 GB limit and be slowed to a speed where synchronous video-based classes and telemedicine are inaccessible for the remainder of the billing period.²⁶¹ Many students would hit the 5 GB limit in less than one week.

Similarly, Verizon's prepaid data plans state in a small screen visible after clicking on "see plan detail" that for the 5 GB and 15 GB plans "[o]nce high-speed data is used (including Mobile Hotspot), you will have 2G speeds the remainder of the month. Your data experience and functionality of some data applications such as streaming video or audio may be impacted unless you purchase a Data Boost."²⁶²

For Verizon's "Unlimited" prepaid plan, clicking through the plan detail link reveals a small screen that states:

Mobile Hotspot and tethering available on the Unlimited Plan for \$5/mo. Includes 10 GB of 5G Nationwide / 4G LTE data, then speeds up to 600 Kbps the remainder of the month. Your data

260. *Simply Prepaid Plans*, T-MOBILE, <https://www.t-mobile.com/support/plans-features/simply-prepaid-plans> (last visited Aug. 17, 2021).

261. See text accompanying *supra* note 220 for data use calculations to access Zoom via a mobile phone using cellular data or as a hot spot.

262. *Prepaid Phone Plans*, VERIZON, https://www.verizon.com/prepaid/?ds_rl=1035790&ds_rl=1275402&cmp=KNC-C-HQ-PRO-R-BP-NONE-Prepaid60back-2K0PX0-PX-BIN-7170000010397525&msclkid=ddc1c137bf8011824eb79f4e00e43911&gclid=CI6ukMfTsu-gCFTuhZQodNVkOOw&gclsrc=ds (last visited Aug. 17, 2021).

experience and functionality of some data applications such as streaming video or audio may be impacted.²⁶³

Similarly for Verizon’s “Unlimited Plus” plan, the clickthrough page explains “[w]hen not in a 5G Ultra Wideband coverage area, Mobile Hotspot access includes 10 GB of 5G Nationwide/4G LTE data, then speeds up to 600 Kbps the remainder of the month.”²⁶⁴

When using a phone to access class or work via videoconferencing, 12.5 hours would consume 10 GB of data. 600 kbps is within the range of 3G speeds that may support Zoom, but synchronous Zoom conferences with multiple speakers using gallery view requires 2.0–4.0 mbps, depending on the number of participants.²⁶⁵ Although Verizon warns that “some data applications such as streaming video or audio may be impacted,”²⁶⁶ many consumers may not understand this precludes telemedicine and video platforms commonly used for school or work.

Videoconferencing such as Zoom conferences used for work, school or telemedicine is distinct from streaming video. Videoconferencing, which “works by connecting two or more computers for direct communication,” is synchronous and can allow screen sharing.²⁶⁷ Streaming video, in contrast, can be recorded or live, but it is controlled by the streamer, allowing people to watch like they would a television show.²⁶⁸

Zoom requires more bandwidth for group video calling (i.e., videoconferencing) than for streaming video. For group video calling Zoom requires “[f]or high-quality video: 1.0 Mbps/600kbps (up/down).”²⁶⁹ For Webinar panelists, Zoom requires 600 kbps (down) for high-quality video.²⁷⁰ A notice that “streaming video or audio may be impacted” may be insufficient

263. *Id.*

264. *Id.*

265. ZOOM, *supra* note 219 (“For gallery view receiving: 2.0Mbps (25 views), 4.0Mbps (49 views)” is required).

266. VERIZON, *supra* note 262.

267. *Video Conferencing vs Live Streaming: What’s Best For Distance Learning?*, VIEW SONIC (Apr. 10, 2020), <https://www.viewsonic.com/library/education/video-conferencing-vs-live-streaming-whats-best-for-distance-learning/>
#:~:text=another%20big%20difference%20is%20the%20ability%20for%20students,could%20be%20a%20greater%20strain%20on%20your%20computer.

268. See Jenny Farver, *Live Streaming vs Video Conferencing—Which should you choose?*, LIGHTSTREAM, <https://golightstream.com/live-streaming-vs-video-conferencing-which-should-you-choose/> (last visited July 6, 2021).

269. *System Requirements For Windows, macOS, and Linux*, ZOOM (Apr. 20, 2021), <https://support.zoom.us/hc/en-us/articles/201362023-system-requirements-for-windows-macos-and-linux>.

270. *Id.*

to alert consumers that videoconferencing could become inaccessible, not merely “impacted” by ISP slowing to 2G speeds.

AT&T’s prepaid data plan states in small print for its 5G and 15G plans, “After 2GB, data speeds are slowed to a maximum of 128Kbps for the rest of the term.”²⁷¹ “AT&T provides no explanation on the offer’s face or in any hyperlinked text regarding what applications or services will not work when users are slowed to 128Kbps for the rest of the term. AT&T does not limit this practice to times of congestion.”²⁷²

AT&T’s prepaid unlimited high-speed data plan says “AT&T may temporarily slow data speeds if the network is busy.”²⁷³ The advertisement does not explain the level of slowness users may experience, how long slowing may last, or how slowdowns will affect user ability to access internet applications (“apps”). For each of those plans, clicking onto “Read the Legal Stuff” reveals a disclosure that states “[a]bility to stream, video resolution, and other data usage (including speed) are not guaranteed, may vary, and be affected by a variety of other factors.”²⁷⁴

FCC rule 47 C.F.R. § 8.1 requires that ISPs disclose “network management practices, performance characteristics, and commercial terms of its broadband [i]nternet access services sufficient to enable consumers to make informed choices regarding the purchase”²⁷⁵ Listing the slowing to 128 kbps or 256 kbps in smaller font without explanation that those speeds will make video-based telemedicine, work, and school impossible is insufficient disclosure to enable consumer choice or protect public safety.

The D.C. Circuit in *Mozilla v. FCC* recognized that the FCC acknowledged that “[t]he competitive process and antitrust would not protect free expression in cases where consumers have decided that they are willing to tolerate some blocking or throttling in order to obtain other things of value.”²⁷⁶ Lack of disclosure about the extent and consequences of throttling undermines consumer choice and the ostensible bargained-for-exchange of values upon which the FCC relied.²⁷⁷

271. *Choose Your AT&T Prepaid Plan*, AT&T, <https://www.att.com/buy/wireless/prepaid/plandetails> (last visited May 24, 2021).

272. BBIC, *Comments on Mozilla Remand*, *supra* note 5, at 47.

273. AT&T, *supra* note 271.

274. *Id.*

275. 47 C.F.R. § 8.1 (2018).

276. *Mozilla Corp. v. FCC*, 940 F.3d 1, 72 (D.C. Cir. 2019).

277. *Cf.* Justin S. Brown & Andrew W. Bagley, *Neutrality 2.0: The Broadband Transition to Transparency*, 25 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 639, 682 (2015) (“The clear disclosure of network management practices and quality of service measures in broadband Internet access providers’ terms of use agreements provides important

Smaller and lighter font, separate clickthrough screens, and lack of explanation of what slowing subscribers to 128 kbps means raises issues about AT&T's compliance with the FTC's permanent injunction in *FTC v. AT&T Mobility*. That injunction prohibits AT&T from "[m]aking any representation about the amount or speed of mobile data, including that the mobile data is unlimited, without disclosing, Clearly and Conspicuously and in Close Proximity to the representation, all Material Restrictions imposed by Defendant."²⁷⁸

Under the FTCA's Deceptive Conduct provisions "placement, proximity, and prominence are key factors for effective disclosure."²⁷⁹ At the dawn of the 21st century, the FTC advised in its Dot.com Disclosure guidelines that "advertisers should consider the *placement* of the disclosure in an ad and its *proximity* to the relevant claim."²⁸⁰

The FTC's Stipulated Injunction with AT&T explained that "Clear[ly] and Conspicuous[ly]" means that "a required disclosure is difficult to miss (i.e., easily noticeable) and easily understandable by ordinary consumers."²⁸¹ It requires that any "visual disclosure, by its size, contrast, location, the length of time it appears, and other characteristics, must stand out from any accompanying text or other visual elements so that it is easily noticed, read, and understood" and be proximate to any "triggering representation."²⁸²

The FCC 2020 RIF Order concluded that concerns that "ISP blocking or throttling that causes harm to public safety are speculative and unlikely to occur," citing the "dearth of real-world examples of public safety harms from blocking or throttling mass market broadband [i]nternet access service."²⁸³ This conclusion rests in part on the FCC's refusal to recognize ISP throttling as a net neutrality violation or a public safety harm. This "linguistic engineering," as Professor Lentz termed the FCC's computer inquiries that created the distinction between information services and common carrier services, has real-world consequences for regulation and service to the public.²⁸⁴

information to consumers to make informed choices about their selection of their high-speed provider and expectations in terms of what they may experience as a subscriber.".)

278. *FTC AT&T Mobility Stipulated Order*, *supra* note 11, at 4, lines 24–26.

279. Sandoval, *Disclosure, Deception, and Deep Packet Inspection*, *supra* note 4, at 667.

280. *Id.* at 667 (citing FTC, DOT COM DISCLOSURES: INFORMATION ABOUT ONLINE ADVERTISING 1 (2000), <https://www.ftc.gov/sites/default/files/attachments/press-releases/ftc-staff-issues-guidelines-internet-advertising/0005dotcomstaffreport.pdf>).

281. *FTC AT&T Mobility Stipulated Order*, *supra* note 11, at 2, lines 26–27.

282. *Id.* at 3, lines 3–5 & 21–23.

283. *2020 RIF Order*, *supra* note 12, at 12358, ¶ 54.

284. Lentz, *supra* note 179, at 443.

Yet, throttling can limit access to safety, economic, educational, health, and other services on the internet. ISP contracts and software maintain and enforce digital exclusion. The Greenlining Institute's report on the digital divide during COVID-19 recounts Felix's story about how easy it is to hit the 3 GB limit and the consequences of speed restrictions once throttling hits:

I have 3 gigs [of smartphone data] a month. It just takes a few hours to run out because I have so many tabs open for all these projects. I tried pairing my cell phone with the computer. But if you nod off for a few hours you lose a gig or so because you're not paying attention. [Then] you have to wait a month for a refresh.²⁸⁵

Financial, credit, and other issues may make other plans infeasible for consumers like Felix. Ambiguity about what activities use 3 GB of data and the consequences of slowing to 2G speeds may steer consumers into choosing plans that appear economical but leave them without functional internet access for many contemporary uses for weeks at a time.

My 2009 article *Disclosure, Deception and Deep-Packet Inspection; The Role of the Federal Trade Commission Act's Deceptive Conduct Prohibitions in the Net Neutrality Debate* argued that "improving ISP disclosure about the extent and breadth of [i]nternet access offered is a necessary but insufficient step to guarantee that the [i]nternet will remain open to all lawful applications."²⁸⁶ Inadequate disclosure in small faded print that does not make the consequences of ISP throttling clear is inconsistent with internet openness and may violate FCC transparency and FTC deceptive conduct laws and regulations. The prevalence of inadequate disclosures across carriers underscores the need for FCC and FTC regulatory action to protect consumers, internet openness, public safety, and the public interest.

VI. RECOMMENDATIONS AND CONCLUSION

To protect consumers, this Article recommends that the FTC review AT&T's compliance with the injunction in *FTC v. AT&T Mobility* and the FTCA's deceptive conduct proscriptions. ISP contract terms are communicated in small letters that fail to define 2G speeds or to fully explain the types of services that will become inaccessible. Such practices raise serious concerns about FTC Act compliance.

285. Vincent Le & Gisella Moya, *On the Wrong Side of the Digital Divide*, GREENLINING INST. (June 2, 2020), <https://greenlining.org/publications/online-resources/2020/on-the-wrong-side-of-the-digital-divide/>.

286. Sandoval, *Disclosure, Deception and Deep-Packet Inspection*, *supra* note 4, at 710–11.

To foster internet openness and protect public safety, the FCC should examine whether ISP disclosures for mobile plans, including prepaid plans and wireline plans, are sufficient under the FCC's transparency rules to enable consumers to make informed choices. This analysis is imperative as lack of adequate disclosure about the extent and effect of ISP throttling undercuts the FCC's EBB program, which relies on ISP disclosure to support consumer choice. AT&T, Verizon, and T-Mobile, along with several other carriers, offered EBB plans.²⁸⁷

The BBIC raised concerns about ISP throttling in its April 2020 comments to the FCC pursuant to the *Mozilla* remand.²⁸⁸ Neither the FCC, the FTC, nor any states have acted since the 2019 *FCC v. AT&T Mobility* settlement to stem ISP throttling that undermines internet openness, public safety, public health, and equity. No party in the FCC's RIF proceedings from 2018–20 offered any “reasonable network management” rationale to defend such throttling of prepaid mobile internet users. Any explanation of technical rationale should be made public to allow examination of the necessity of such practices and reasonable alternatives.

This Article recommends the FCC and states, including state public utility commissions with jurisdiction over ISPs, collect data on the extent to which ISPs are slowing users to speeds that make modern applications such as mapping and videoconferencing fail. Neither the FCC nor the FTC require ISPs to publicly report on the frequency and duration of throttling to 2G speeds. Many consumers are unable to reliably access common applications demanded for modern work, school, and health access, and they may not realize that their ISP is throttling them or understand why it occurred. The FCC may require carriers subject to FCC transparency rules to provide data about throttling practices pursuant to its authority under Communications Act § 409(e) to access books and records of entities subject to FCC jurisdiction.²⁸⁹

Nor is data publicly available about the race, ethnicity, or gender of those most affected by ISP throttling practices. Follow-up studies to the Student Home Internet Connectivity report should investigate whether students are subject to ISP-induced slowdowns and, if so, for how long, and the effect of that throttling on education.

287. *Emergency Broadband Benefit Providers*, FCC, <https://www.fcc.gov/emergency-broadband-benefit-providers#California> (last visited May 25, 2021).

288. BBIC, *Comments on Mozilla Remand*, *supra* note 5, at 14, 44–49.

289. See 47 C.F.R. § 0.111(h) (authorizing the FCC per 47 U.S.C. § 409(e) to inspect books, records, contracts, agreements, and document subject to any investigation).

The FCC must address ISP throttling practices that undermine access to education, health, economic resources and opportunities, and social resources.²⁹⁰ Concomitantly, the FCC should reassess its definition of “advanced service.” The 25/3 mbps standard falls gigabits short of current needs and fails to lay the groundwork for a robust and equitable internet access. ISP throttling compromises robust internet access as envisioned in FCC advanced service definitions. The FCC should also examine ISP throttling in its section 706 reports on advanced telecommunications services and in its analysis of the communications marketplace.

Following major ISPs “written commitments not to engage in practices considered to violate open [i]nternet principles, including blocking and throttling,”²⁹¹ ISPs should abandon throttling to 2G speeds that undermine major internet, health, safety, educational, and economic uses. Consistent with corporate pledges to promote equity and inclusion,²⁹² ending ISP practices that close the digital schoolhouse, healthcare, and economic opportunity door by throttling users back to the 90s would enable equity, inclusion, public health, and public safety.

The FCC is charged with a statutory duty to promote the “safety of life and property through the use of wire and radio [wireless] communications.”²⁹³ Fulfilling this duty demands consideration of the broad range of internet uses and users that affect and promote safety of life and property.

The FCC’s public safety and advanced services analysis must put the public at the center of public safety and public interest regulation. Doing so requires the FCC to recognize and safeguard internet-based communications within and between the public, civic organizations, and critical infrastructure sectors as critical to public safety and democracy. The FCC and FTC can begin this process by taking steps to stop ISP throttling that endangers public health, safety, education and undermines opportunity.

290. See BBIC, *Comments on Mozilla Remand*, *supra* note 5, at 19–20.

291. 2020 RIF Order, *supra* note 12, at 12356, ¶ 50.

292. See, e.g., *Diversity and Inclusion*, AT&T, <https://about.att.com/pages/diversity> (last visited, May 26, 2021); *Diversity and Inclusion*, VERIZON, <https://www.verizon.com/about/our-company/diversity-and-inclusion> (last visited May 26, 2021); *Diversity, Equity, and Inclusion (DEI) at T-Mobile*, T-MOBILE, https://www.t-mobile.com/news/_admin/uploads/2020/08/816819_DEI-Factsheet.pdf.

293. 47 U.S.C. § 151; see also *Mozilla v. FCC*, 940 F.3d 1, 59 (D.C. Cir. 2019).

FROM LEX INFORMATICA TO THE CONTROL REVOLUTION

Julie E. Cohen[†]

ABSTRACT

Legal scholarship on the encounter between networked digital technologies and law has focused principally on how law should respond to new technological developments and has spent much less time considering what that encounter might signify for the shape of legal institutions themselves. This essay focuses on the latter question. Within fields like technology studies, labor history, and economic sociology, there is a well-developed tradition of studying the ways that new information technologies and the “control revolution” they enabled—in brief, a quantum leap in the capacity for highly granular oversight and management—have elicited long-term, enduring changes in the structure and operation of economic organizations. I begin by considering some lessons of work in that tradition for law understood as a set of organizations constituted for the purpose of governance. Next, I turn the lens inward, offering some observations about techlaw scholarship that are essentially therapeutic. The disruptions of organizational change have affected scholars who teach, think, and write about techlaw in ways more profound than are commonly acknowledged and discussed. It seems fitting, in a symposium dedicated to Joel Reidenberg’s life and work, to use the process of grief as a device for exploring the arc of techlaw scholarship over its first quarter century. The fit is surprisingly good and the takeaways relatively clear: if, as I intend to suggest, the organizational forms that underpin our familiar legal institutions have been in the process of evolving out from under us, we still have choices to make about how legal institutions optimized for the information economy will be constituted. Finally, I identify two sets of important considerations that should inform processes of organizational and institutional redesign.

DOI: <https://doi.org/10.15779/Z38C53F239>

© 2021 Julie E. Cohen.

[†] Mark Claster Mamolen Professor of Law and Technology, Georgetown Law. My thanks to Lindsey Barrett, Yochai Benkler, Hannah Bloch-Wehba, Ryan Calo, Erin Carroll, April Falcon Doss, evelyn douek, Neil Richards, Pamela Samuelson, Rebecca Tushnet, Rory Van Loo, Ari Waldman, Rebecca Wexler, Lauren Willis, and Jonathan Zittrain for their helpful comments, to Joel Reidenberg for leading the way, and to Christina Wing for research assistance.

TABLE OF CONTENTS

I.	INTRODUCTION	1018
II.	THE CONTROL REVOLUTION IN GOVERNANCE	1021
III.	GRIEF COUNSELING FOR LAW PROFESSORS	1027
	A. BACK(ING IN) TO THE FUTURE	1028
	B. THE WRATH OF NETWORKS	1031
	C. GETTING TO MEH.....	1035
	D. THE UNBEARABLE LIGHTNESS OF DEVOLUTION.....	1038
	E. SO NOW WHAT?	1042
IV.	THE RULE OF LAW AFTER THE CONTROL REVOLUTION	1043
	A. NETWORKED GEOGRAPHIES: MAPPING FLOWS, CONTROL POINTS AND FAILURE MODES.....	1043
	B. LEGAL NORMATIVITIES: COUNTERING SYSTEMATIC ABUSES OF POWER	1046
V.	CONCLUSION: WWJD?	1049

I. INTRODUCTION

In the beginning (techlaw-wise) came two texts.¹ Together, they defined an agenda for exploring the encounter between networked digital technologies and law—and together, they also encoded methodological fractures and disciplinary blind spots that persist today. “Lex Informatica” was an article published by a legal scholar—Joel Reidenberg, to whose memory this symposium is dedicated—for an audience of other legal scholars.² Complex and subtle, it explored the ways government authorities might reassert themselves within pathways and processes defined in the first instance by computer networks and digital code. The other text—Lawrence Lessig’s *Code and Other Laws of Cyberspace*—began as a law review article but evolved into a book crafted for a more general audience.³ Punchy and attention-grabbing, it offered a simple, flat taxonomy of regulatory forces, each assertedly different in kind and origin from the others, and identified ways that processes emerging

1. This Article borrows the useful neologism coined by Rebecca Crootof & BJ Ard, *Structuring Techlaw*, 34 HARV. J.L. & TECH. 348 (2021).

2. Joel R. Reidenberg, *Lex Informatica: The Formulation of Information Policy Rules Through Technology*, 76 TEX. L. REV. 553 (1998).

3. LAWRENCE LESSIG, *CODE AND OTHER LAWS OF CYBERSPACE* (1999); Lawrence Lessig, *The Law of the Horse: What Cyberlaw Might Teach*, 113 HARV. L. REV. 501 (1999).

in the domain of digital code might frustrate other processes traditionally located in the domain of law.

If one accepted the premise that governing new technological activities required new types of responses from law- and policymakers,⁴ the two texts dictated different approaches to identifying those responses. Consider, for example, the question of what (if anything) to do about copyright management technologies designed to enable licensing but simultaneously frustrating other important copyright policy goals. Or consider the question of what (if anything) to do about a global networked communications architecture that promised to evade governance by both nation-states and other intermediaries traditionally entrusted with ensuring information quality. Because “Lex Informatica” was the product of a mind trained in both North American and European ways of thinking about law and regulation, it turned automatically to the mechanics of injecting regulatory authority into the processes by which networked communications technologies and associated standards were being developed and deployed.⁵ Because *Code* was a product of the “New Chicago School,” it foregrounded the influence of markets and norms and the bottom-up solutions they might generate.⁶ There are layers upon layers of irony here. *Code*, but not “Lex Informatica,” purported to offer a new approach to theorizing the regulatory properties of technology; “Lex Informatica” was more pragmatic in its orientation. And yet “Lex Informatica,” but not *Code*, surfaced the complex *interplay* between regulatory forces. “Lex Informatica” framed new digital formations as situated opportunities for interventions by policymakers and other interested actors—an approach broadly compatible with decades of accumulated, interdisciplinary learning on emergent sociotechnical processes—whereas *Code* described an elemental regulatory struggle that unfolded as a contest over *terra nullius* and that resonated with the reigning neoliberal ethos of the era.

Gradually but inexorably, however, new developments began to pose questions that the two texts did not contemplate—questions about what the encounter between networked digital technologies and law might signify for the shape of legal institutions themselves. For example, what does it mean to require ongoing “compliance” with a remedial decree directed to the operation of data-driven, algorithmic processes? What corrective actions can remedial

4. Some did not. The canonical example is Frank H. Easterbrook, *Cyberspace and the Law of the Horse*, 1996 U. CHI. LEGAL F. 207 (1996), though I hesitate to give it yet another citation.

5. Reidenberg, *supra* note 2, at 583–92.

6. LESSIG, *supra* note 3, at 85–99, 122–41, 164–85, 223–26. See generally Lawrence Lessig, *The New Chicago School*, 27 J. LEGAL STUD. 661 (1998); Mark Tushnet, “Everything Old Is New Again”: Early Reflections on the “New Chicago School”, 1998 WISC. L. REV. 579 (1998).

orders directed to smaller actors within networked information ecosystems plausibly require, and what obligations should be placed on the larger actors that design and operate such systems? What organizational configurations and practices are needed to ensure sufficient public accountability of compliance operations? What organizational configurations and practices are needed to ensure that data-driven surveillance processes designed to operate on populations afford sufficient dignity and respect to individuals and communities?

As these examples suggest, there is an important difference between understanding networked digital technologies as “regulating” in ways that might challenge or complement law and understanding such technologies as catalyzing deep structural transformation in organizations of all sorts, including the organizational forms of legal institutions carefully stewarded—and venerated—over decades and centuries. This essay takes the latter perspective as its point of departure. Within fields like technology studies, labor history, and economic sociology, there is a well-developed tradition of studying the ways that new information technologies and the “control revolution” they enabled—in brief, a quantum leap in the capacity for highly granular oversight and management—have elicited long-term, enduring changes in the structure and operation of economic organizations.⁷ Part II considers some lessons of work in that legal tradition for law understood as a set of organizations constituted for the purpose of governance.

Part III turns the lens inward, offering some observations about techlaw scholarship that are essentially therapeutic. The disruptions of organizational change have affected scholars who teach, think, and write about techlaw in ways more profound than are commonly acknowledged and discussed. It seems fitting, in a symposium dedicated to Joel Reidenberg’s life and work, to use the process of grief as a device for exploring the arc of techlaw scholarship over its first quarter century. The fit is surprisingly good and the takeaways relatively clear: if, as I intend to suggest, the organizational forms that underpin our familiar legal institutions have been in the process of evolving out from

7. The term comes from JAMES R. BENIGER, *THE CONTROL REVOLUTION: TECHNOLOGICAL AND ECONOMIC ORIGINS OF THE INFORMATION SOCIETY* (1986). It was later appropriated by ANDREW SHAPIRO, *THE CONTROL REVOLUTION: HOW THE INTERNET IS PUTTING INDIVIDUALS IN CHARGE AND CHANGING THE WORLD WE KNOW* (1999), which misunderstood the nature of the shift in control that digital networks represented. For an important but conceptually distinct exploration of the evolving role of control within digital information and communications networks, see LAURA DENARDIS, *THE INTERNET IN EVERYTHING: FREEDOM AND SECURITY IN A WORLD WITH NO OFF SWITCH* (2020) (arguing that digital networks are undergoing a phase shift from communication to control as their principal purpose).

under us, we still have choices to make about how legal institutions optimized for the information economy will be constituted. Learning to identify the reflex reactions emanating from grief's intermediate stages will help us make better choices.

Building on the insights from Parts II and III, Part IV identifies two sets of important considerations that should inform the redesign of legal institutions after the control revolution. One set of considerations involves efficacy. Legal institutions for the control revolution require organizational forms that are optimized to networked information and communication geographies, flows, points of control, and failure modes. The second relates to normative sufficiency; the redesign of legal institutions requires appropriately framed rule-of-law criteria. Part V concludes.

II. THE CONTROL REVOLUTION IN GOVERNANCE

The relationship between law and networked digital technologies is, and always has been, a two-way street. Legal actors respond to new technological developments, but the principals in new technological dramas also exploit and work to reconfigure legal and governance regimes in ways that are most congenial to their own activities and goals.⁸ Scholarship in the law and society tradition has long acknowledged and grappled with the power of self-interested advocacy to reshape the *rules* by which litigants, regulated industries, and other actors in legal dramas must play.⁹ But “law” also consists of *organizations* constituted for the purpose of governance, and those organizations also are affected by sociotechnical change. Additionally, although some institutional realignments reflect the intentional efforts of self-interested *actors*, sociotechnical transformation produces both intended and unintended *systemic effects*. Here, I bring classic mid-twentieth-century studies of the encounter between for-profit organizations and emerging informational capabilities to bear on law's evolving organizational and systemic accommodations to the informational era.

First, some important definitions: By “organization,” I mean to refer to an entity constituted to achieve a particular goal, with sets of rules that define its

8. See generally JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM (2019). On the concept of technological drama, see Bryan Pfaffenberger, *Technological Dramas*, 17 SCI., TECH., & HUM. VALUES 282 (1992).

9. See generally Marc Galanter, *Why the “Haves” Come Out Ahead: Speculations on the Limits of Legal Change*, 9 L. & SOC’Y REV. 95 (1974); MORTON J. HORWITZ, THE TRANSFORMATION OF AMERICAN LAW, 1780–1860 (1977).

structure and govern the practices of its members.¹⁰ For-profit companies like Amazon or Microsoft are organizations; courts, regulatory agencies, and industry standards bodies are organizations, too. By “institution,” I mean either an organization or an otherwise well-defined set of practices that serves a public or social purpose.¹¹ Some institutions, such as courts and administrative agencies or the pre-internet “press,” have (or had) distinct organizational forms. Others, such as contract law and tort law, do not, but even institutions of the latter sort (which I will call rule-based institutions) may reflect assumptions about the particular organizational contexts within which they will be interpreted and applied. This is particularly true for rule-based institutions, such as corporate law or administrative law, that are intended to provide structural specifications for organizations. Finally, by “legal institution,” I mean an institution whose outputs are constituted as binding by political authority. In this essay, I will be concerned with transformations in the organizational forms (or, for rule-based institutions, the assumed organizational contexts) of legal institutions.

As scholars in fields like technology studies, labor history, and economic sociology began to study the organizational impacts of new information technologies, they noticed that organizations undergo profound changes as new methods of seeing and managing their own activities are taken on board. In his magisterial study of organizational transformation, historian of technology James Beniger gave this process a name—the “control revolution”—that is equally useful for thinking about changes in the organization of governance.¹² Scholarly accounts of the interrelationships between information technologies and the organization of economic production have three more particular lessons for legal scholars—including not only those who say they study techlaw but also those who insist that they don’t and won’t.

The first lesson of the control revolution is that it changes *how* organizations produce outputs. As Beniger showed, the control revolution in production involved radical jumps in the quantity and granularity of information generated by newly mechanized production processes and correspondingly radical changes in the configuration of control processes for

10. See generally Saylor Breckenridge & Scott Savage, *Organizations*, OXFORD BIBLIOGRAPHIES (July 27, 2011), <https://www.oxfordbibliographies.com/view/document/obo-9780199756384/obo-9780199756384-0039.xml>.

11. See generally Fabio Rojas, *Institutions*, OXFORD BIBLIOGRAPHIES (Aug. 26, 2013), <https://www.oxfordbibliographies.com/view/document/obo-9780199756384/obo-9780199756384-0132.xml>.

12. BENIGER, *supra* note 7.

collecting and managing information and directing production accordingly.¹³ New information technologies afforded perspectives on production that were simultaneously panoptic and synoptic. One could zoom in on a particular set of operations in a highly granular way—for example, by investigating the relationship between a particular machine configuration and production throughput. One could also zoom out for a large-scale view of the organization's operations—for example, asking and answering questions about geographic and seasonal variation in demand. And, increasingly, one could ask new types of questions about the interplay between the granular and the systemic—for example, questions about how workspace configurations, supply chains, and a host of other factors might be rearranged to respond most effectively to serve and reinforce new patterns of mass production and consumption.¹⁴

The second lesson of the control revolution is that it changes *what* organizations produce. Newly granular and comprehensive control of production logistics enabled organizations to formulate new production plans that would enable them to capitalize on the infrastructural and informational investments they were making. So, for example, as it became possible to manage food production and distribution over extended geographic areas, the nature of food production changed to emphasize pre-processing, standardized packaging, and distribution via self-service supermarkets.¹⁵

The third lesson of the control revolution is that changes in the *how* and *what* of production created points of entry for changing ideologies about what and whom production was (good) for—about *why* organizations produce. As labor historians like Harry Braverman and Sanford Jacoby showed, the control revolution facilitated large-scale changes in the conditions of labor. New cadres of managerial workers were needed to operate the new systems for communication and control, and the managerial turn in the organization of production aligned with the goals of those wishing to shift control of production away from workers and concentrate it among the owners of capital.¹⁶ The control revolution and the managerial turn in the organization of production unfolded alongside other, technologically-mediated transformations in financial markets—in particular, the emergence of

13. *See id.* Beniger rejected rigidly deterministic explanations for these changes, indicating that he viewed them as coupled in varying degrees of tightness with other economic and social developments. *See id.* at 6–10.

14. *Id.* at 293–317.

15. *See id.* at 248–78, 337–42.

16. *See* HARRY BRAVERMAN, LABOR AND MONOPOLY CAPITAL 251–69 (Monthly Rev. Press 1998) (1974); SANFORD JACOBY, EMPLOYING BUREAUCRACY: MANAGERS, UNIONS, AND THE TRANSFORMATION OF WORK IN THE 20TH CENTURY (rev. ed. 2004).

increasingly complex and financialized performance metrics and investment vehicles—and these developments also reinforced the growing power of management and capital more generally.¹⁷ As the twentieth century wore on, the cumulative effects of those changes proved congenial to a neoliberal worldview that envisioned government as existing principally to steward and validate the results of market processes.¹⁸ For all of these historically contingent reasons, the instrumentalities of the control revolution in economic production increasingly were directed toward surplus extraction for the benefit of managers and investors.

In retrospect, it seems utterly naïve to have thought that these lessons would not apply to the organizational forms of legal institutions. Consider a few examples:

In the domain of dispute resolution, networked information technologies and systems have facilitated widespread outsourcing of small, low-dollar value disputes in areas such as consumer satisfaction and human resources, and they also have enabled parties to large-scale tort and regulatory litigation to develop new organizational mechanisms for producing and managing settlements in ways only nominally under supervision by courts.¹⁹ Changes in how disputes are resolved have shaped what dispute resolution produces. Both large-scale settlements and privatized processes for resolving small-scale disputes require and normalize elaborate sets of managerial practices for administering payments and (sometimes) for measuring and documenting compliance with agreed organizational changes. These processes require new cadres of managerial workers and may also involve an assortment of other third parties—compliance auditors, litigation financiers, specialized arbitrators and

17. See GIOVANNI ARRIGHI, *THE LONG TWENTIETH CENTURY: MONEY, POWER, AND THE ORIGINS OF OUR TIMES* (new and updated ed. 2010); GRETA KRIPPNER, *CAPITALIZING ON CRISIS: THE POLITICAL ORIGINS OF THE RISE OF FINANCE* (2011); Natascha Van der Zwan, *Making Sense of Financialization*, 12 SOCIO-ECON. REV. 99 (2014).

18. See THE ROAD FROM MONT PELERIN: THE MAKING OF THE NEOLIBERAL THOUGHT COLLECTIVE (Philip Mirowski & Dieter Plehwe eds., paperback ed. 2015); Nicholas Gane, *The Governmentalities of Neoliberalism: Panopticism, Post-Panopticism, and Beyond*, 60 SOCIO. REV. 611, 627–29 (2012); Gerard Hanlon, *The First Neo-Liberal Science: Management and Neo-Liberalism*, 52 SOCIO. 298 (2018).

19. See COHEN, *supra* note 8, at 155–67. On dispute outsourcing, see Lauren B. Edelman & Mark Suchman, *When the "Haves" Hold Court: Speculations on the Organizational Internalization of Law*, 33 L. & SOC'Y REV. 941 (1999); Rory Van Loo, *The Corporation as Courthouse*, 33 YALE J. ON REG. 547 (2016). On the flexible production of large-scale dispute resolution, see Abbe R. Gluck & Elizabeth Chamblee Burch, *MDL Revolution*, 96 N.Y.U. L. REV. 1 (2021); David M. Jaros & Adam S. Zimmerman, *Judging Aggregate Settlement*, 94 WASH. U. L. REV. 545 (2017); Daniel J. Solove & Woodrow Hartzog, *The FTC and the New Common Law of Privacy*, 114 COLUM. L. REV. 583 (2014).

judges, and so on.²⁰ Their outputs typically do not consist of citable opinions articulating rule-based formulations about proper conduct and appropriate liability. These developments have elicited both criticism and praise; within the legal academy, there is contestation over what and whom dispute resolution is (good) for.²¹

In the administrative state, regulators charged with overseeing the operations of the informational economy must demand, evaluate, and act on new kinds of representations by regulated entities. In fields ranging from finance and healthcare to pollution control and avionics, data-driven algorithmic processes demand correspondingly sophisticated oversight mechanisms.²² Meanwhile, regulators who administer large-scale benefits and revenue systems rely ever more heavily on automated tools for case management and decision-making.²³ Changes in how regulators exercise their oversight and decision-making authority translate into changes in both the form and the substance of regulatory outputs. Guidances, collaborative best practices statements, and consent decrees requiring changes to regularized control practices increasingly stand in for more formal rules and more definite enforcement orders. Decisions about technical standard-setting and information systems procurement play increasingly central roles, and as regulated activities in sectors such as banking, consumer finance, environmental protection, and the like have grown ever more informationally complex, the regulatory landscape has widened to include a diverse group of third-party auditors, systems vendors, and other compliance intermediaries.²⁴

20. See COHEN, *supra* note 8, at 159–64; Charles F. Sabel & William H. Simon, *Destabilization Rights: How Public Law Litigation Succeeds*, 117 HARV. L. REV. 1016 (2004).

21. See generally GILLIAN K. HADFIELD, *RULES FOR A FLAT WORLD: WHY HUMANS INVENTED LAW AND HOW TO REINVENT IT FOR A COMPLEX ECONOMY* (2017); Brooke D. Coleman, *One Percent Procedure*, 91 WASH. L. REV. 1005 (2016); Jason Parkin, *Aging Injunctions and the Legacy of Institutional Reform Litigation*, 70 VAND. L. REV. 167 (2017); Judith Resnik, *Diffusing Disputes: The Public in the Private of Arbitration, the Private in Courts, and the Erasure of Rights*, 124 YALE L.J. 2804 (2015); Sabel & Simon, *supra* note 20.

22. See COHEN, *supra* note 8, at 170–200. See generally Mehrsa Baradaran, *Regulation by Hypothetical*, 67 VAND. L. REV. 1247 (2014); Cary Coglianese, *The Limits of Performance-Based Regulation*, 50 U. MICH. J. L. REFORM 525 (2017); Kathryn Judge, *Fragmentation Nodes: A Study in Financial Innovation, Complexity, and Systemic Risk*, 64 STAN. L. REV. 657 (2012); Wendy Wagner & Martin Murillo, *Is the Administrative State Ready for Big Data?*, KNIGHT FIRST AMEND. INST. (Apr. 30, 2021), <https://knightcolumbia.org/content/is-the-administrative-state-ready-for-big-data>; Karen Yeung, *Algorithmic Regulation: A Critical Interrogation*, 12 REG. & GOV. 505 (2018).

23. See Ryan Calo & Danielle Keats Citron, *The Automated Administrative State: A Crisis of Legitimacy*, 70 EMORY L.J. 797 (2021).

24. See COHEN, *supra* note 8, at 189–93; Kenneth A. Bamberger, *Technologies of Compliance: Risk and Regulation in a Digital Age*, 88 TEX. L. REV. 669 (2010); Kenneth A. Bamberger,

Concurrently, the universe of theoretical accounts of regulatory behavior has widened to include new narratives about the virtues of regulatory devolution and enlightened self-governance, and these narratives too are contested.²⁵ In particular, the proliferation of new regulatory inputs, outputs, and intermediaries does not seem to be producing more effective oversight of information-economy activities, and data-driven algorithmic processes also enable new types of gaming that can be difficult to detect.²⁶

Within law enforcement agencies and inside the national security state, networked information technologies have facilitated new data-driven surveillance practices focused on the ready availability of digital traces of human movement and communication that can be gathered remotely without tasking individual officers to follow suspects, execute search warrants, and tap phone lines. That tectonic shift in how surveillance is conducted has elicited, and worked to naturalize, new tools and capabilities—for computer forensic investigations, for collecting and analyzing digital images, for gathering and correlating location information, and for conducting data-driven predictive analysis—and those tools and capabilities in turn generate new outputs that must be evaluated.²⁷ Some data-driven surveillance processes have elicited new types of managerial oversight, but others continue to operate in ways

Regulation as Delegation: Private Firms, Decisionmaking, and Accountability in the Administrative State, 56 DUKE L.J. 377 (2006); Deirdre K. Mulligan & Kenneth A. Bamberger, *Procurement as Policy: Administrative Process for Machine Learning*, 34 BERKELEY TECH. L.J. 773 (2019).

25. See generally Jody Freeman, *The Private Role in Public Governance*, 75 N.Y.U. L. REV. 543 (2000); Michael Livermore & Richard L. Revesz, *Regulatory Review, Capture, and Agency Inaction*, 101 GEO. L.J. 1337 (2012); Orly Lobel, *The Renew Deal: The Fall of Regulation and the Rise of Governance in Contemporary Legal Thought*, 89 MINN. L. REV. 342 (2004); Jodi L. Short, *The Paranoid Style in Regulatory Reform*, 63 HASTINGS L.J. 633 (2012).

26. See, e.g., Cary Coglianese & Jennifer Nash, *The Law of the Test: Performance-Based Regulation and Diesel Emissions Control*, 34 YALE J. ON REG. 33 (2017); Makena Kelly, *FCC's Net Neutrality Rollback Overwhelmed by Bogus Industry Comments, Investigation Finds*, THE VERGE (May 6, 2021), <https://www.theverge.com/2021/5/6/22422818/net-neutrality-rollback-ajit-pai-telecom-broadband-new-york-attorney-general>; STAFF OF S. COMM. ON ANTITRUST, COM. AND ADMIN. L. COMM. ON JUDICIARY, 116TH CONG., INVESTIGATION OF COMPETITION IN DIGITAL MARKETS (2020).

27. See, e.g., SARAH BRAYNE, PREDICT AND SURVEIL: DATA, DISCRETION, AND THE FUTURE OF POLICING (2020); Orin S. Kerr, *Searches and Seizures in a Digital World*, 119 HARV. L. REV. 531 (2005); Eric Lichtblau, *More Demands on Cell Carriers in Surveillance*, N.Y. TIMES (July 8, 2012), <https://www.nytimes.com/2012/07/09/us/cell-carriers-see-uptick-in-requests-to-aid-surveillance.html>; Andrea Roth, *Machine Testimony*, 126 YALE L.J. 1972 (2017); CLARE GARVIE, ALVARO BEDOYA & JONATHAN FRANKLE, THE PERPETUAL LINEUP: UNREGULATED POLICE FACE RECOGNITION IN AMERICA, GEORGETOWN L.: CTR. ON PRIV. & TECH. (Oct. 18, 2016), <https://www.perpetuallineup.org/>.

seemingly unconstrained by existing constitutional and statutory protections.²⁸ Many cross preexisting organizational lines, emerging out of procurement processes, hybrid public-private partnerships, and cross-jurisdictional policing and border enforcement initiatives.²⁹ Many have come to seem uniquely unaccountable to the broader public whose interests they are supposed to be serving.

The ultimate lesson of the control revolution for law is that networked information technologies are not simply new modes of knowledge production to be governed, but also powerful catalysts for organizational restructuring that change the enterprise of governance (and so, necessarily, also that of law³⁰) from the inside out. They produce new organizational formations that resemble the idealized legal-institutional models taught in law school courses only vestigially and incidentally. And the new organizational formations of the control revolution generate outputs that familiar modes of legal-institutional understanding cannot parse.

III. GRIEF COUNSELING FOR LAW PROFESSORS

For legal scholars, large-scale, disruptive change in the organizational forms of legal institutions is not an abstraction to be studied at arms-length. It represents a profound loss that reverberates through every facet of our carefully burnished, collective professional identity. As we move from teaching students about institutional configurations notable chiefly because they no longer exist outside the pages of casebooks, to writing about those same configurations in the pages of law reviews as though they still deserved to

28. On managerial oversight, see, e.g., Daphna Renan, *The FISC's Stealth Administrative Law*, in GLOBAL INTELLIGENCE OVERSIGHT: GOVERNING SECURITY IN THE TWENTY-FIRST CENTURY 121 (Zachary K. Goldman & Samuel J. Rascoff ed., 2016). On unconstrained surveillance behavior, see, e.g., Kimberly N. Brown, *Outsourcing, Data Insourcing, and the Irrelevant Constitution*, 49 GA. L. REV. 607 (2015); Sara Morrison, *A Surprising Number of Government Agencies Buy Cellphone Location Data. Lawmakers Want to Know Why*, VOX RECODE (Dec. 2, 2020), <https://www.vox.com/recode/22038383/dhs-cbp-investigation-cellphone-data-brokers-venntel>.

29. See, e.g., Caroline Haskins, *Scars, Tattoos, and Licenses Plates: This is What Palantir and the LAPD Know About You*, BUZZFEED NEWS (Sept. 29, 2020), <https://www.buzzfeednews.com/article/carolinehaskins1/training-documents-palantir-lapd>; Elizabeth E. Joh, *The Undue Influence of Surveillance Technology Vendors on Policing*, 92 N.Y.U. L. REV. ONLINE 101 (2017); Mulligan & Bamberger, *supra* note 24; Priscilla M. Regan, Torin Monahan & Krista Craven, *Constructing the Suspicious: Data Production, Circulation, and Interpretation by DHS Fusion Centers*, 47 ADMIN. & SOC'Y 74 (2015); Morgan Simon, *Investing in Immigrant Surveillance: Palantir and the #NoTechForICE Campaign*, FORBES (Jan. 15, 2020), <https://www.forbes.com/sites/morgansimon/2020/01/15/investing-in-immigrant-surveillance-palantir-and-the-notechforice-campaign/?sh=3ea7ca7b7707>.

30. For more on this distinction, see *infra* Part IV.B.

command the lion's share of our attention and energy, to envisioning the possible futures of a system of governance whose central tenets and operational presumptions no longer seem to cohere, both our day-to-day routines and our more sustained intellectual projects continually remind us that the system into which we were trained has lost its moorings.

Put differently and more starkly, we experience grief—and grief calls for a type of introspection to which the legal academy is unaccustomed. Here, I use Kubler-Ross's well-known five-stage framework as a device for mapping scholarly responses to the control revolution's disruptions.³¹ (Without question, this sort of exercise is reductive and risks oversimplification. Even so, it can be useful for diagnostic purposes. My aim here is to prompt reflection, not to urge outright dismissal of important works that fall into each of these categories.)

A. BACK(ING IN) TO THE FUTURE

The first response to large-scale, disruptive, and profoundly grief-inducing change tends to be denial. So too within techlaw scholarship. An essential mode of legal theorizing about networked digital technologies has been the assertion that nothing *really fundamental* about legal subject x has changed, will change, or should change as a result of development y .

Denial is a tricky subject to unpack because the rearview mirror represents law's methodological wheelhouse (and the objects reflected in it are always much closer than they appear). Judges and legislators alike move forward only slowly and tentatively, continually looking backward, identifying analogies, and redeploying familiar common law concepts—even when interpreting statutes clearly intended to craft new institutional settlements. But they also must contend with the entrepreneurialism of practicing lawyers and the self-interested actors they represent. Sometimes, however, denial benefits powerful actors, and the ensuing dynamic represents law's most dangerous endemic failure mode: a downward spiral into institutional paralysis catalyzed by self-interest, self-importance, and conceptual rigidity.

There is no better illustration of the dynamic of denial spiraling into institutional paralysis than the path charted by the mainstream of scholarship and advocacy about the First Amendment implications of the networked information revolution. Consider the debates about “deplatforming” unwanted speakers. For some First Amendment traditionalists, questions

31. ELISABETH KUBLER-ROSS & DAVID KESSLER, ON GRIEF AND GRIEVING: FINDING THE MEANING OF GRIEF THROUGH THE FIVE STAGES OF LOSS (2005) (identifying as the five stages denial, anger, bargaining, depression, and acceptance, and exploring their complexities and interrelationships).

about the power to deplatform are easy to answer because of the public-private distinction that (in their view, appropriately) structures the universe of speech protections.³² The earliest scholarly commentary on deplatforming worried about private power to stifle dissenting speech emanating from members of marginalized groups and from the political left. For those commentators, there were equally traditionalist answers: the “back to the future” strategies of treating platforms as either public forums or company towns obligated to permit speech with which they disagree.³³ More recently, deplatforming efforts directed at purveyors of white supremacist and other hate speech has caused advocates of the “company town” approach to reconsider that position, even as avowed traditionalists from the right float the very different “back to the future” strategy of subjecting platforms to common carrier obligations.³⁴ Others, meanwhile, have fallen back on a different type of traditionalist argument: faith in the “marketplace of ideas” to produce clear rejection of white supremacy and hate once brought into the light of day.³⁵

The problem, though, is that none of these arguments reckons adequately with underlying transformations in the structure of speech environments. The “long tail” marketplace of the platform-mediated speech environment, which facilitates access to and monetization of all perspectives, does not seem to be furthering large-scale rejection of white supremacy, ethnonationalism, and hate. Rather, it is nurturing them. Although mainstream media organizations

32. See, e.g., Eric Goldman, *Of Course the First Amendment Protects Google and Facebook (and It's Not a Close Question)*, KNIGHT FIRST AMEND. INST. (Mar. 23, 2021), <https://knightcolumbia.org/content/course-first-amendment-protects-google-and-facebook-and-its-not-close-question>.

33. See, e.g., Steven G. Gey, *Reopening the Public Forum—From Sidewalks to Cyberspace*, 58 OHIO ST. L.J. 1535 (1998); Dawn C. Nunziato, *The Death of the Public Forum in Cyberspace*, 20 BERKELEY TECH. L.J. 1115 (2005); see also Kyle Langvardt, *Regulating Online Content Moderation*, 106 GEO. L.J. 1353 (2018) (invoking the company town doctrine to justify public oversight of content moderation practices).

34. *Biden v. Knight First Amend. Inst.* At Columbia Univ., 141 S. Ct. 1220, 1222–26 (2021) (Thomas, J., concurring); Eugene Volokh, *Trump Was Kicked Off Twitter. Who's Next?*, N.Y. TIMES (Jan. 11, 2021) <https://www.nytimes.com/2021/01/11/opinion/trump-twitter-facebook-parler.html>; Eugene Volokh, *Facebook “Removing Content Containing the Phrase ‘Stop the Steal’”*, REASON (Jan. 11, 2021) <https://reason.com/volokh/2021/01/11/facebook-removing-content-containing-the-phrase-stop-the-steal/>.

35. See, e.g., Danielle Keats Citron & Helen Norton, *Intermediaries and Hate Speech: Fostering Digital Citizenship for Our Information Age*, 91 B.U. L. REV. 1435, 1474 (2011) (“By challenging hate speech with counter-speech, intermediaries can help transform online dialogue by documenting the continuing existence of racism and other forms of hatred while concomitantly rebutting it.”); Richard Delgado & Jean Stefancic, *Hate Speech in Cyberspace*, 49 WAKE FOREST L. REV. 319, 341 (2014) (“Denouncing the group or individual publicly can demonstrate to users of the Internet that disseminating hate through this medium brings consequences and can give pause to others who might be tempted to follow suit.”).

and prominent political figures continue to function as principal content hubs in networked media ecosystems, platform-based techniques for profiling users and their social networks and for routing, upranking, and recommending content have been game-changers. These techniques have supplied powerful, flexible tools for seeding mainstream media environments with disinformation, hate, and polarizing discursive frames, recruiting new adherents to hate-based worldviews, and expanding extremist communities and networks.³⁶ Because the platform-mediated speech environment relies on probabilistic profiles and engagement metrics to route, uprank, and recommend content and communities, post hoc content- and speaker-level interventions do not meaningfully disrupt the mechanisms by which extremist sentiment diffuses across interlinked networks.³⁷

The blunt Newtonian instruments supplied by current First Amendment doctrine are wholly inadequate to the task of apportioning governance authority within such spaces. Because current doctrine presumes functioning speech markets populated by rational listeners, it assumes away the distinctive dysfunctions of platform-based information environments optimized for behavioral microtargeting, automatic engagement, and rapid, cascading spread.³⁸ And modes of constitutional argumentation that simply reassert private authority to govern such processes in a more fine-grained way

36. On the centrality of mainstream media organizations and political figures, see YOCHAI BENKLER, ROBERT FARIS & HAL ROBERTS, NETWORK PROPAGANDA: MANIPULATION, DISINFORMATION, AND RADICALIZATION IN AMERICAN POLITICS (2018); YOCHAI BENKLER, CASEY TILTON, BRUCE ETLING, HAL ROBERTS, JUSTIN CLARK, ROBERT FARIS, JONAS KAISER & CAROLYN SCHMITT, MAIL-IN VOTER FRAUD: ANATOMY OF A DISINFORMATION CAMPAIGN (Berkman Klein Ctr. Research Publication No. 2020-6, 2020), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3703701. On the complementary, amplifying effects of platform-based media infrastructures, see Joan Donovan, *Source Hacking: Media Manipulation in Practice*, DATA & SOC'Y (Sept. 4, 2019), <https://datasociety.net/library/source-hacking-media-manipulation-in-practice/>; Anthony Nadler, Matthew Crain & Joan Donovan, *Weaponizing the Digital Influence Machine*, DATA & SOC'Y (Oct. 17, 2018), <https://datasociety.net/library/weaponizing-the-digital-influence-machine/>; Manoel Horta Ribeiro, Raphael Ottoni, Robert West, Virgilio A. F. Almeida & Wagner Meira, *Auditing Radicalization Pathways on YouTube*, 2020 PROC. 2020 CONF. ON FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY 131 (Jan. 2020), <https://arxiv.org/pdf/1908.08313.pdf>; Francesca Tripodi, *Searching for Alternative Facts*, DATA & SOC'Y (May 16, 2018), <https://datasociety.net/library/searching-for-alternative-facts/>.

37. See, e.g., Corin Faife & Dara Kerr, *Facebook Said it Would Stop Recommending Anti-Vaccine Groups. It Didn't*, THE MARKUP (May 20, 2021), <https://themarkup.org/citizen-browser/2021/05/20/facebook-said-it-would-stop-recommending-anti-vaccine-groups-it-didnt>; Nadler et al., *supra* note 36; Tripodi, *supra* note 36.

38. See Julie E. Cohen, *Tailoring Election Regulation: The Platform is the Frame*, 4 GEO. L. TECH. REV. 641, 642–55 (2020).

undermine efforts to render the increasingly complex manifestations of platform power publicly accountable.

My argument here is not about the way that ostensibly neutral moves within free speech discourse work systematically to benefit already-powerful economic and political actors and to effect erasure of other distinctions that really do matter (in part because I think that is so clearly true as to be beyond serious debate); rather, I want to underscore a more basic point. It is long past time to call into question interpretive conventions devised during the era of broadcast media for a constitutional text that is itself an artifact of an even earlier era, to acknowledge and retire the assumptions about information flow that have continued to inform those interpretive conventions even when they no longer describe reality, and to pursue other ways of honoring the foundational commitments the text sought to express.³⁹

The costs of denial are existential. Failure to recognize and reckon with the paradigm shifts in our information environment may yet herald the end of both our particular 250-year experiment with democratic self-governance and other democratic experiments worldwide. Fortunately, the therapeutic lens also suggests that First Amendment denialism represents an evolutionary stage that techlaw scholarship and our legal system more broadly may yet transcend.

B. THE WRATH OF NETWORKS

The second stage of grief is anger, and here an initial caveat is in order. I do not mean to use the stages-of-grief device to diminish techlaw scholarship expressing anger at the ways in which new forms of informationalized power have mobilized legal institutions to work systemic economic and racialized injustice.⁴⁰ Righteous wrath over law's complicity in the perpetuation of systemic injustice has a centrally important role in legal scholarship and public

39. See generally Tim Wu, *Is the First Amendment Obsolete?*, 117 MICH. L. REV. 547 (2018); Lawrence Lessig, *Fidelity in Translation*, 71 TEX. L. REV. 1165 (1993); cf. Danielle Keats Citron & Neil M. Richards, *Four Principles for Digital Expression (You Won't Believe #3!)*, 95 WASH. L. REV. 1353 (2018) (discussing non-constitutional strategies for translating concerns about expressive liberty into online environments).

40. See, e.g., VIRGINIA EUBANKS, *AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR* (2018); Ifeoma Ajunwa, *Race, Labor, and the Future of Work*, in OXFORD HANDBOOK OF RACE AND LAW (Devon Carbado, Emily Houh & Khiara M. Bridges eds., 2020); Alvaro M. Bedoya, *Privacy as a Civil Right*, 50 N.M. L. REV. 301 (2020); Rashida Richardson, *Government Data Practices as Necropolitics and Racial Arithmetic*, GLOBALDATAJUSTICE (Oct. 8, 2020), <https://globaldatajustice.org/covid-19/necropolitics-racial-arithmetic>; see also RUHA BENJAMIN, *RACE AFTER TECHNOLOGY: ABOLITIONIST TOOLS FOR THE NEW JIM CODE* (2019); SAFIYA UMOJA NOBLE, *ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM* (2018); CATHY O'NEIL, *WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY* (2016).

interest advocacy. The anger that I want to spotlight here is different and more unique to techlaw. It is the anger of the frustrated (cyber)libertarian who takes issue with the asserted need to have a system of law at all.

Confronted with the increasing inadequacy and imperfection of traditional governance mechanisms, some legal scholars began to advance variations on the theme of frustrated utopianism. They argued that centralized gatekeeping was the enemy, that bottom-up creativity and crowd-sourced ordering were potent forces for good, and that under such circumstances, law's highest and best goal was to minimize its own footprint.⁴¹

Anger and frustrated utopianism have been especially notable features of scholarly and policy debates about the future of digital copyright. Copyright law has always represented an effort to balance the competing goals of commercial reward and creative and expressive freedom. Because networked digital environments enable both new types of freedom and new types of control—and because the major industry stakeholders had long been accustomed to dictating the shape of new legislation—proposals for digital-era copyright legislation were highly contentious.⁴² As the major copyright industries pushed for legal recognition of expanded control and the mainstream of copyright scholarship resisted proposals that seemed overly draconian, the perfect became the enemy of the good. Some scholars rejected compromises that would entail any sacrifice of flexibility to copy, manipulate, or share digital content. So, for example, proposals for automated filtering of content uploaded to file-sharing platforms drew criticism because filtering algorithms could not duplicate the flexibility and nuance that fair use doctrine required, and proposals to restrict copying of audio and/or video files were criticized on the ground that depriving users of the ability to reuse content would limit their creative freedom.⁴³ Complaints about emergent linking and

41. See, e.g., David R. Johnson & David Post, *Law and Borders—The Rise of Law in Cyberspace*, 48 STAN. L. REV. 1367 (1996); LAWRENCE LESSIG, *THE FUTURE OF IDEAS: THE FATE OF THE COMMONS IN A CONNECTED WORLD* (2002); SHAPIRO, *supra* note 7; see also Jane R. Bambauer & Derek E. Bambauer, *Information Libertarianism*, 105 CALIF. L. REV. 335 (2017). Two prominent works in cyberutopian canon were more nuanced, offering accounts of bottom-up creativity that were also keenly sensitive to the prospects for abuses of private power and to the roles that law might play in constraining such abuses. See YOCHAI BENKLER, *THE WEALTH OF NETWORKS: HOW SOCIAL PRODUCTION TRANSFORMS MARKETS AND FREEDOM* (2007); JONATHAN ZITTRAIN, *THE FUTURE OF THE INTERNET AND HOW TO STOP IT* (2008).

42. For good summaries, see BILL HERMAN, *THE FIGHT OVER DIGITAL RIGHTS: THE POLITICS OF COPYRIGHT AND TECHNOLOGY* (2013); JESSICA LITMAN, *DIGITAL COPYRIGHT* (2006).

43. See, e.g., Dan L. Burk & Julie E. Cohen, *Fair Use Infrastructure for Copyright Management Systems*, 15 HARV. J.L. & TECH. 41 (2001); Rebecca Tushnet, *Copy This Essay: How Fair Use*

embedding practices that channeled advertising revenues away from legacy content producers to new digital intermediaries were roundly mocked as the last gasps of industrial-economy gatekeepers seeking to silence new citizen performers, documentarians, and journalists.⁴⁴

Having spent some quality time in this stage of scholarly grief myself, I continue to think that some of the anger at copyright overreach was and is amply justified—as we are about to see, compromise requires two sides—but it also has delayed a much-needed reckoning with the governance challenges of networked digital environments.⁴⁵ And because power abhors a vacuum, legislative and policy stalemates over the legitimate reach of copyright law have privileged narrower, self-interested arrangements that reinforce economic power. The leading copyright intermediaries have retained and in some cases expanded their traditional strongholds, while newer information platforms have emerged as the default aggregators for new forms of cultural production (such as short video clips) and for self-published content.⁴⁶ These institutional settlements have not been costless. Platform intermediaries have moved quickly to design their own automated filtering systems and develop their own

Doctrine Harms Free Speech and How Copying Serves It, 114 YALE L.J. 535, 558–60 (2004); Rebecca Tushnet, *I Put You There: User-Generated Content and Anti-Circumvention*, 12 VAND. J. ENT. & TECH. L. 889 (2010). For more recent variations on these themes, see Dan L. Burk, *Algorithmic Fair Use*, 86 U. CHI. L. REV. 283 (2019); Rebecca Tushnet, *All of This Has Happened Before and All of It Will Happen Again: Innovations in Copyright Licensing*, 29 BERKELEY TECH. L.J. 1447 (2014) [hereinafter Tushnet, *All of This Has Happened Before*].

44. See, e.g., Dan Hunter & Gregory F. Lastowka, *Amateur-to-Amateur*, 46 WM. & MARY L. REV. 951 (2005). For more recent variations on this theme, see Annemarie Bridy, *The Price of Closing the “Value Gap”: How the Music Industry Hacked EU Copyright Reform*, 22 VAND. J. ENT. & TECH. L. 323 (2020); Mike Masnick, *The Bizarre Reaction To Facebook’s Decision To Get Out Of The News Business In Australia*, TECHDIRT (Feb. 18, 2021), <https://www.techdirt.com/articles/20210217/22383446265/bizarre-reaction-to-facebooks-decision-to-get-out-news-business-australia.shtml>; Calla Wahlquist, *Australia’s Proposed Media Code Could Break the World Wide Web, Says the Man who Invented It*, THE GUARDIAN (Jan. 19, 2021), <https://www.theguardian.com/media/2021/jan/20/australias-proposed-media-code-could-break-the-world-wide-web-says-the-man-who-invented-it>.

45. See, e.g., Burk & Cohen, *supra* note 43; Julie E. Cohen, *Pervasively Distributed Copyright Enforcement*, 95 GEO. L.J. 1 (2006); Julie E. Cohen, *Copyright and the Perfect Curve*, 53 VAND. L. REV. 1799 (2000); Julie E. Cohen, *Copyright and the Jurisprudence of Self-Help*, 13 BERKELEY TECH. L.J. 1089 (1998); Julie E. Cohen, Lochner in Cyberspace: *The New Economic Orthodoxy of “Rights Management”*, 97 MICH. L. REV. 462 (1998).

46. See, e.g., Sara Morrison, *Why Facebook Banned (and then Unbanned) News in Australia*, VOX RECODE (Feb. 25, 2021), <https://www.vox.com/recode/22287971/australia-facebook-news-ban-google-money>; THOMAS POELL, DAVID B. NIEBORG & BROOKE ERIN DUFFY, *THE PLATFORMIZATION OF CULTURAL PRODUCTION* (1st ed. 2021); see also Guy Pessach, *Beyond IP—The Cost of Free: Informational Capitalism in a Post IP Era*, 54 OSGOOD HALL L. REV. 225, 239–45 (2016) (arguing that networked platform ecosystems create new patterns of media concentration and content standardization).

linking and embedding conventions, and those choices in turn have systematically shifted creative agency away from human beings and digital advertising revenues away from entities such as news providers that serve important public needs.⁴⁷

The push to elevate generativity over gatekeeping also produced section 230 of the Communications Decency Act (CDA 230), which insulates information intermediaries from most forms of civil liability for most expressive choices by their users, while granting them broad latitude to engage in content moderation operations of their own design.⁴⁸ Over the years, debates about the wisdom of the institutional settlement reflected in CDA 230 have demonstrated the continuing influence of First Amendment denialism (thereby illustrating that grief is not a linear journey). CDA 230's supporters and advocates have expressed both cyberlibertarian outrage at the prospect of imposing gatekeeping obligations on the new digital frontier and backward-looking lawyerly confidence in the crude public-private distinction that the statute encodes.⁴⁹ They have held to both positions even as the death of gatekeeping has demonstrated more and more powerfully that generativity is a scalar, not a vector; that torrents of xenophobia, hate, and conspiracy theory are generative in all the wrong ways; and that platforms govern their own operations continually in ways that amplify those torrents because they are profitable.

I do not mean to be glib about the urgency of the threats to freedom of expression surfaced by cyberlibertarian legal scholarship. Conflicts between institutional control and expressive freedom arise in any centrally governed

47. On automated filtering and its effects, see Ira Steven Nathenson, *Civil Procedures for a World of Shared and User-Generated Content*, 48 U. LOUISVILLE L. REV. 911, 937–38 (2010); Sonia K. Katyal, *Filtering, Piracy Surveillance and Disobedience*, 32 COLUM. J.L. & ARTS 401, 412–13 (2008). On proprietary linking and embedding conventions, see Tushnet, *All of This Has Happened Before*, *supra* note 43. On the impacts of platformization on journalism, see MIKE ANANNY, NETWORKED PRESS FREEDOM: CREATING INFRASTRUCTURES FOR A PUBLIC RIGHT TO HEAR (2018); Erin C. Carroll, *Platforms and the Fall of the Fourth Estate*, 78 MD. L. REV. 529 (2019).

48. 47 U.S.C. § 230(c)(1)–(2).

49. See, e.g., Eric Goldman, *Why Section 230 is Better than the First Amendment*, 95 NOTRE DAME L. REV. ONLINE 33 (2019); Jeff Kosseff, *Defending Section 230: The Value of Intermediary Immunity*, 15 J. TECH. L. & POL'Y 123 (2010). For representative work advocating reform of section 230, see Danielle Keats Citron & Benjamin Wittes, *The Internet will Not Break: Denying Bad Samaritans Sec. 230 Immunity*, 86 FORDHAM L. REV. 401 (2018); see also Mary Anne Franks, Mike Goodwin, Jeff Kosseff & Andrés Martínez, *Where Do We Go From Here with Section 230*, SLATE (Dec. 15, 2020), <https://slate.com/technology/2020/12/legal-scholars-mary-anne-franks-mike-goodwin-and-jeff-kosseff-on-section-230-of-the-cda.html> (recounting a roundtable discussion among three prominent commentators with different perspectives on section 230 reform).

regime, but in networked spaces they are both endemic and especially difficult to resolve. Sublimated anger about law's inherent repressiveness, however, is untenable as a long-term survival strategy both for the legal academy and more generally for any moderately complex society. Concededly, governance institutions are always-already imperfect and freedom-limiting, and they also must fight continual rearguard actions against capture, abuse, and overreach. But they are also necessary.

C. GETTING TO MEH

After anger comes bargaining—another wheelhouse mode for lawyers. In the particular context of techlaw scholarship, bargaining expresses hope that the control revolution's disruptions might be accommodated by making relatively minor tweaks and adjustments to the law's core institutions and routines.

In many contexts, bargaining is an ordinary and expected way of producing good-enough results for all parties—and sometimes, it can yield creative resolutions vastly superior to the remedies that a court would be empowered to devise.⁵⁰ But bargaining presumes both a relatively equal distribution of bargaining power and a clear understanding of the universe of effective interventions. If one party is relatively well-resourced and well-equipped to undertake costly litigation, it will have little incentive to agree to concessions that seem unnecessary. If the same party also controls access to information about feasible remedial actions and need not share that information, it may be impossible for the other party to know what interventions to propose.⁵¹

The ongoing debates about content moderation and digital privacy protection illustrate the perils of bargaining without discernible leverage. In the United States, the potent combination of statutory immunity for content moderation operations, broad privilege to harvest and process most user personal information, and sheer economic might has allowed the dominant platform firms to assume that refusal to compromise is costless. Whether defying requests for information from regulators, violating issued enforcement orders, or deflecting questions from members of Congress, their behavior has manifested clear awareness of their own impunity.⁵² Additionally, continuing a

50. See generally Thomas O. Main, *ADR: The New Equity*, 74 U. CIN. L. REV. 329 (2005) (describing the potential flexibility of ADR).

51. See generally Russell Korobkin, *A Positive Theory of Legal Negotiation*, 88 GEO. L.J. 1789, 1804-06 (2000) (analyzing the effects of information costs in ADR).

52. See U.S. SENATE COMM. ON COMMERCE, SCIENCE, AND TRANSPORTATION, OFFICE OF OVERSIGHT AND INVESTIGATIONS MAJORITY STAFF, *A REVIEW OF THE DATA BROKER INDUSTRY: COLLECTION, USE, AND SALE OF CONSUMER DATA FOR MARKETING PURPOSES* 10–11 (Dec. 18, 2013); U.S. FED. TRADE COMM'N, *DATA BROKERS: A CALL FOR*

theme developed in the previous two Sections, information technology firms of all sizes have leveraged the staying power of First Amendment denialism, weaponizing arguments about expressive liberty in the service of a more narrowly self-interested vision of immunity from regulatory oversight. Fueled by their contributions, a litigation campaign to extend First Amendment protection to all forms of information processing has been gathering strength.⁵³

The content moderation and digital privacy debates also illustrate the costs of insufficient access to relevant information about the operation of data-driven algorithmic processes. Technology firms—especially the dominant platforms that wield the greatest economic and cultural power—go to extraordinary lengths to keep their processes for profiling users and routing content shrouded in secrecy.⁵⁴ Additionally, although platforms govern their own operations continually, they share only the most basic and superficial information about how internal governance processes work.⁵⁵ Without such information, it is impossible to formulate concrete proposals for governing differently.

As a different illustration of the costs of insufficient access to relevant information, consider the evolving discussions about algorithmic fairness, accountability, and transparency in search, digital advertising, and image recognition. As awareness of endemic problems of bias in automatic activity began to spread, journalists and scholars documented troubling patterns and

TRANSPARENCY AND ACCOUNTABILITY 7–10 (2014); *Transcript of Mark Zuckerberg's Senate Hearing*, WASH. POST (Apr. 10, 2018), <https://www.washingtonpost.com/news/the-switch/wp/2018/04/10/transcript-of-mark-zuckerbergs-senate-hearing/>; *Transcript of Zuckerberg's Appearance before House Committee*, WASH. POST (Apr. 11, 2018), <https://www.washingtonpost.com/news/the-switch/wp/2018/04/11/transcript-of-zuckerbergs-appearance-before-house-committee/>; SHOSHANA ZUBOFF, *THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER* 145–48, 159–61 (2019). See generally FRANK PASQUALE, *THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION* (2015).

53. See *Sorrell v. IMS Health, Inc.*, 564 U.S. 552, 563–70 (2011); Jane R. Bambauer, *Is Data Speech?*, 66 STAN. L. REV. 57 (2014); Eugene Volokh & Donald Falk, *First Amendment Protection of Search Engine Search Results*, 8 J. L. ECON. & POL'Y 88 (2012); see generally COHEN, *supra* note 8, at 89–97; Amanda Shanor, *The New Lochner*, 2016 WISC. L. REV. 133 (2016); Jameel Jaffer & Ramya Krishnan, *Clearview AI's First Amendment Theory Threatens Privacy—and Free Speech Too*, SLATE (Nov. 17, 2020), <https://slate.com/technology/2020/11/clearview-ai-first-amendment-illinois-lawsuit.html>.

54. See sources in *supra* note 52.

55. See evelyn douek, *The Rise of Content Cartels*, KNIGHT FIRST AMEND. INST. (Feb. 11, 2020), <https://knightcolumbia.org/content/the-rise-of-content-cartels>. See generally Rory Van Loo, *The Missing Regulatory State: Monitoring Businesses in an Age of Surveillance*, 72 VAND. L. REV. 1563 (2019).

advocacy organizations filed discrimination lawsuits.⁵⁶ Particularly when considered in light of the longer history of antidiscrimination litigation, the research and the lawsuits seemed to dictate fairly obvious corrective measures—adjust the algorithms to make them fairer, exclude particularly problematic data fields (such as race or gender), and so on. In a series of highly publicized statements and settlements, digital giants such as Facebook and Google agreed to make those sorts of changes.⁵⁷ In reality, however, such limited interventions only make problems of bias more intractable. Machine learning algorithms reproduce and reinforce the patterns that exist in the data sets used to train them. Even when first-order data about protected characteristics such as race or gender is placed off limits, they will detect and reproduce preexisting patterns of systemic racial or gender-based disadvantage.⁵⁸ Addressing patterns of injustice that are fundamentally social,

56. *See, e.g.*, Dep't of Hous. & Urban Dev. v. Facebook, Inc., FHEO No. 01-18-0323-8 (2019) (enforcement action against Facebook for discriminatory housing advertisements); Nat'l Fair Hous. All. v. Facebook, Inc., No. 1:18-cv-02689 (S.D.N.Y. Feb. 6, 2019) (lawsuit against Facebook for discriminatory housing advertisements); Divino Group LLC v. Google LLC, No. 5:19-cv-04749 (N.D. Cal. Aug. 13, 2019) (lawsuit against YouTube and its parent company Google for algorithmic discrimination of creators based on race and sexual orientation); Alistair Barr, *Google Mistakenly Tags Black People as 'Gorillas,' Showing Limits of Algorithms*, WALL ST. J. (July 1, 2015), <https://www.wsj.com/articles/BL-DGB-42522>; James Hale, *Four Black Creators File Suit Against YouTube, Alleging Racial Discrimination in Algorithm*, TUBEFILTER (June 19, 2020), <https://www.tubefilter.com/2020/06/19/youtube-black-creators-lawsuit-algorithm-discrimination/>; Latanya Sweeney, *Discrimination in Online Ad Delivery*, COMM. ACM, May 2013, <http://cacm.acm.org/magazines/2013/5/163753-discrimination-in-online-ad-delivery/>.

57. *See, e.g.*, Sheryl Sandberg, *Doing More to Protect Against Discrimination in Housing, Employment and Credit Advertising*, FACEBOOK (Mar. 19, 2019), https://about.fb.com/news/2019/03/protecting-against-discrimination-in-ads/?ref=FBBlog_UpdatesToHousing; *Updates to Housing, Employment and Credit Ads in Ads Manager*, FACEBOOK (Aug. 26, 2019), <https://www.facebook.com/business/news/updates-to-housing-employment-and-credit-ads-in-ads-manager>; Tom Simonite, *When It Comes to Gorillas, Google Photos Remains Blind*, WIRED (Jan. 11, 2018), <https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/>.

58. *See, e.g.*, Ava Kofman & Ariana Tobin, *Facebook Ads Can Still Discriminate Against Women and Older Workers, Despite a Civil Rights Settlement*, PROPUBLICA (Dec. 13, 2019), <https://www.propublica.org/article/facebook-ads-can-still-discriminate-against-women-and-older-workers-despite-a-civil-rights-settlement>; Jeremy B. Merrill, *Does Facebook Still Sell Discriminatory Ads?*, THE MARKUP (Aug. 25, 2020), <https://themarkup.org/ask-the-markup/2020/08/25/does-facebook-still-sell-discriminatory-ads>; Piotr Sapiezynski, Avijit Ghosh, Levi Kaplan, Alan Mislove & Aaron Rieke, *Algorithms that "Don't See Color": Comparing Biases in Lookalike and Special Ad Audiences*, (arXiv, Working Paper No.1912.07579, 2019), <https://arxiv.org/pdf/1912.07579.pdf>; Simonite, *supra* note 57. *See generally* David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn about Machine Learning*, 51 U.C. DAVIS L. REV. 65 (2017); Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 86 FORDHAM L. REV. 1085 (2018).

not technical, requires different types of intervention in the design of data-driven algorithmic processes—and ensuring efficacy requires the ability to audit those processes, even when their operators would prefer to keep them proprietary. Generally speaking, the dominant technology firms have resisted granting the sorts of access that would enable researchers to hold them accountable, and some have taken aggressive steps to block researchers from collecting such information on their own.⁵⁹

Unlike denial and anger, bargaining under conditions that guarantee failure has some salutary uses. It underscores power disparities, and it highlights the information deficits that prevent good-faith negotiation and foreclose mutually acceptable compromise. Unless those pathologies can be addressed, however, that is all it is good for. Bargaining remains, at best, a way station in the process of reckoning with bereavement.

D. THE UNBEARABLE LIGHTNESS OF DEVOLUTION

The fourth stage of grief is depression. Those who have been bereaved begin to acknowledge a future indelibly stamped with loss but remain unable to envision anything other than emptiness and absence. Like anger in scholarly work about techlaw, depression is often sublimated. Unlike scholarly anger, which finds its outlet in rejection of imposed legal constraints, scholarly depression masquerades as cheerful optimism about law's increasing marginality. Techlaw scholarship in the depressive mode frames the initial, powerfully self-interested governance formations that have begun to emerge—often, governance formations resulting from the lopsided bargaining described above—as both inevitable and inevitably beneficial.

The wish to put a bright face on corporate performances of accountability did not begin with techlaw; rather, it is broadly reflective of the devolution of governance in an era of ascendant neoliberalism and extractive capitalism. As private economic power increasingly has succeeded at placing itself beyond the reach of law both domestically and globally, lawyers and policymakers have fallen back on optimistic exhortations about corporate social responsibility,

59. See Issie Lapowsky, *Platforms vs. PhDs: How Tech Giants Court and Crush the People Who Study Them*, PROTOCOL (Mar. 19, 2021), <https://www.protocol.com/nyu-facebook-researchers-scraping>. See generally Thomas Kadri, *Digital Gatekeepers*, 99 TEX. L. REV. 951 (2020). A notable exception, so far, is The Markup's Citizen Browser Project. See, e.g., Alfred Ng & Corin Faife, *Facebook Pledges to Remove Discriminatory Credit and Loan Ads Discovered by The Markup*, THE MARKUP (May 4, 2021), <https://themarkup.org/citizen-browser/2021/05/04/facebook-pledges-to-remove-discriminatory-credit-and-loan-ads-discovered-by-the-markup>.

often set forth as nonbinding statements of “principles” and “best practices” designed to serve as fulcrum points for assertions of moral authority.⁶⁰

The very earliest developments in the policy discourse around online content moderation followed this pattern. The Global Network Initiative, a voluntary association of global information businesses formed in 2008, promulgated principles that were intended to empower its members to resist authoritarian states’ demands for censorship, and the United Nations released a series of special reports on the protection of fundamental human rights in networked digital environments.⁶¹

But depressive celebrations of private authority over content moderation also have attached themselves to more concrete forms of privatized governance. The 2018 Santa Clara Principles for Accountability and Transparency in Content Moderation set forth recommendations that included publication of statistics about complaint resolution and provision of notice and appeal rights.⁶² Many social media companies have adopted the recommendations, and some have gone further, publishing information about their criteria for complaint resolution.⁶³

Some legal scholars and tech activists have celebrated these developments while downplaying the fact that the transparency afforded into private content governance *operations* remains relatively low and that public accountability for the design of such systems—and of the content recommendation systems that are their operational counterparts—is essentially nonexistent.⁶⁴ Some embrace

60. See John Ruggie (Representative of the Secretary-General on the Issue of Human Rights and Transnational Corporations and Other Business Enterprises), *Guiding Principles on Business and Human Rights: Implementing the United Nations “Protect, Respect and Remedy” Framework*, Human Rights Council, U.N. Doc. A/HRC/17/31 (Mar. 21, 2011); *The Ten Principles of the UN Global Compact*, UNITED NATIONS GLOB. COMPACT, <https://www.unglobalcompact.org/what-is-gc/mission/principles> (last visited Jan. 20, 2022).

61. *The GNI Principles*, GLOB. NETWORK INITIATIVE, <https://globalnetworkinitiative.org/gni-principles/>. For the most recent United Nations Report, see David Kaye, Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Human Rights Council, U.N. Doc. A/HRC/38/35 (Apr. 6, 2018), <https://www.undocs.org/A/HRC/38/35>; Michael Blowfield & Jędrzej George Frynas, *Setting New Agendas: Critical Perspectives on Corporate Social Responsibility in the Developing World*, 81 INT’L AFF. 499 (2005).

62. ACCESS NOW ET AL., *The Santa Clara Principles on Transparency and Accountability in Content Moderation*, <https://santaclaraprinciples.org/>.

63. See, e.g., Casey Newton, *Facebook Makes its Community Guidelines Public and Introduces an Appeals Process*, THE VERGE (Apr. 24, 2018), <https://www.theverge.com/2018/4/24/17270910/facebook-community-guidelines-appeals-process>.

64. See, e.g., Hannah Bloch-Wehba, *Global Platform Governance: Private Power in the Shadow of the State*, 72 SMU L. REV. 27, 72–74 (2019); Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598 (2018); Jillian C. York, *The Santa*

the flowering of private governance in ways that also reflect the lingering influences of denial and anger; for these authors, the assertion of public governance authority is unlikely to improve matters and is far more likely to introduce unacceptable risks.⁶⁵

Both reactions ensure that new ventures in private governance are heralded even as other possible institutional settlements remain underexplored. So, for example, the Facebook Oversight Board, a body of legal and human rights experts constituted by Facebook (on the advice of an eminent Harvard constitutional law professor) to undertake “review” of selected content moderation decisions, has received breathless coverage in the media and praise from some academic commentators, even though it has very little actual authority.⁶⁶ It accepts very few cases and can recommend action only on the particular content that is before it. It is not charged to recommend more sweeping changes to Facebook’s content moderation policies and practices. It also lacks authority to review the policies and practices that drive content *amplification* or the processes by which Facebook recommends its Groups and

Clara Principles During COVID-19: More Important Than Ever, ELEC. FRONTIER FOUND. (May 11, 2020), <https://www.eff.org/deeplinks/2020/05/santa-clara-principles-during-covid-19-more-important-ever>; see also Marianna B. Ganapini, Camylle Lanteigne & Abhishek Gupta, REPORT PREPARED BY THE MONTREAL AI ETHICS INSTITUTE FOR THE SANTA CLARA PRINCIPLES FOR CONTENT MODERATION (2020), <https://arxiv.org/pdf/2007.00700.pdf> (characterizing the principles as valuable baseline obligations and identifying additional needs); Spandana Singh & Leila Doty, *The Transparency Report Tracking Tool: How Internet Platforms are Reporting on the Enforcement of Their Content Rules*, NEW AMERICA (Apr. 8, 2021), <https://www.newamerica.org/oti/reports/transparency-report-tracking-tool/>; Daphne Keller & Paddy Leerssen, *Facts and Where to Find Them*, in SOCIAL MEDIA AND DEMOCRACY: THE STATE OF THE FIELD, PROPSECTS FOR REFORM 220 (Nathaniel Persily & Joseph A. Tucker ed., 2020); Nicolas P. Suzor, Sarah Myers West, Andrew Quodling & Jillian York, *What do we Mean When we Talk About Transparency? Toward Meaningful Transparency in Commercial Content Moderation*, 13 INT’L J. COMM’CN 1517 (2019).

65. See, e.g., Bloch-Wehba, *supra* note 64; Eric Goldman, *Content Moderation Remedies*, 28 MICH. TECH. L. REV. 1 (2021); DAPHNE KELLER, INTERNET PLATFORMS: OBSERVATIONS ON SPEECH, DANGER, AND MONEY (Hoover Working Grp. Nat’l Sec., Tech., & L. Aegis Series Paper No. 1807, 2018), https://www.hoover.org/sites/default/files/research/docs/keller_webreadypdf_final.pdf.

66. See, e.g., Kate Klonick, *Inside the Making of Facebook’s Supreme Court*, THE NEW YORKER (Feb. 12, 2021), <https://www.newyorker.com/tech/annals-of-technology/inside-the-making-of-facebooks-supreme-court>; Kate Klonick, *The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression*, 129 YALE L. J. 2418 (2020); Steven Levy, *Why Mark Zuckerberg’s Oversight Board May Kill His Political Ad Policy*, WIRED (Jan. 28, 2020), <https://www.wired.com/story/facebook-oversight-board-bylaws/>; Mark Sullivan, *Exclusive: The Harvard Professor Behind the Facebook Oversight Board Defends Its Role*, FAST COMPANY (July 8, 2019), <https://www.fastcompany.com/90373102/exclusive-the-harvard-professor-behind-facebooks-oversight-board-defends-its-role>.

Pages.⁶⁷ Scholarly explorations of the prospects for systematic, public oversight of the data-driven, algorithmic processes that amplify hate and disinformation remain relatively rare.⁶⁸

The growing body of literature about the processes and mechanisms of privacy governance supplies additional examples of techlaw scholarship in the depressive mode. As described in Part II, the two decades-long push to institute appropriate oversight of collection, processing, and use of personal information has produced vast new compliance industries dedicated to the pursuit, perfection, and legitimation of self-governance. As Ari Waldman documents, the processes of compliance are performative—they have as both their clear purpose and their undeniable effect the legitimation of existing practices that serve tech industry interests while allowing individuals very little informational self-determination and regulators very little direct authority to shape industry behavior.⁶⁹ Many privacy scholars put a bright face on these developments, using terms like “coregulation” and “collaborative governance” to describe practices that involve almost no collaboration and produce even less accountability.⁷⁰ Others praise the California Consumer Privacy Act (CCPA) for its boldness, choosing not to dwell on the fact that the CCPA’s principal governance mechanism—post hoc individual assertion of control

67. See evelyn douek, “What Kind of Oversight Board Have You Given Us?,” 2020 U. CHI. L. REV. ONLINE 1 (2020); Rebecca MacKinnon, *The Facebook Oversight Board Did the Best it Could on the Trump Decision*, SLATE (May 5, 2021), <https://slate.com/technology/2021/05/trump-facebook-ban-ruling-oversight-board-power.html>; Siva Vaidhyanathan, *Facebook and the Folly of Self-Regulation*, WIRED (May 9, 2020), <https://www.wired.com/story/facebook-and-the-folly-of-self-regulation/>; see also evelyn douek, *Facebook’s Oversight Board: Move Fast with Stable Infrastructure and Humility*, 21 N.C. J.L. & TECH. 1 (2019) (attempting to define a more realistic and constructive path for the board’s possible interventions).

68. For preliminary discussions, see Cohen, *supra* note 38; evelyn douek, *Verified Accountability*, THE HOOVER INST. (Sept. 17, 2019), <https://www.hoover.org/research/verified-accountability>; evelyn douek, *Second Wave Content Moderation Institutional Design: From Rights to Regulatory Thinking* (Jan. 10, 2022) [hereinafter douek, *Second Wave Content Moderation*], https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4005326.

69. See generally Ari Ezra Waldman, *Privacy, Practice, and Performance*, 110 CALIF. L. REV. (forthcoming 2021); Ari Ezra Waldman, *Privacy Law’s False Promise*, 97 WASH. U. L. REV. 773 (2020).

70. See, e.g., KENNETH A. BAMBERGER & DEIRDRE K. MULLIGAN, *PRIVACY ON THE GROUND: DRIVING CORPORATE BEHAVIOR IN THE UNITED STATES AND EUROPE* (2015); Dennis D. Hirsch, *The Law and Policy of Online Privacy: Regulation Self-Regulation, or Co-Regulation?*, 34 SEATTLE U. L. REV. 439, 465–66 (2011); Margot Kaminski, *Binary Governance: Lessons from the GDPR’s Approach to Algorithmic Accountability*, 92 S. CALIF. L. REV. 1529, 1557–77 (2019).

rights—largely reinforces private authority over the mechanisms and patterns of data collection and use.⁷¹

To be fair, the urge to sublimate depression about the shortcomings of privatized governance solutions also reflects the fact that many other approaches to governing information-economy activities seem to be worse. In particular, authoritarian states have developed a suite of strategies for weaponizing social media, coercing platform compliance with content removal mandates, and extending and enforcing surveillance mandates, and those strategies are manifestly antidemocratic.⁷² If the only alternative to private ordering is authoritarianism, private ordering doesn't seem so bad. That proposition, however, tends to be assumed rather than proved, and it has become a favorite tech industry talking point. U.S.-based information technology firms have worked to position global governance debates as zero-sum games in which the reigning U.S. deregulatory ethos is the only serious alternative to authoritarian rule more broadly.⁷³ The result has been a steady downward spiral toward a future in which effective *democratic* governance of the control revolution seems increasingly out of reach.

E. SO NOW WHAT?

After depression comes acceptance. Unlike depression, however, acceptance does not simply entail resignation to a “new normal” consisting of continuing absence. It also represents an opportunity for new beginnings. If the legacy organizational models underpinning legal institutions have changed beyond recognition, making it infeasible simply to reassert their continuing primacy, perhaps it is time to envision new ones. And if the new, largely privatized governance formations emerging at the intersection of law and technology are not producing the sorts of governance that we want or need, perhaps other kinds of change are now in order. Part IV suggests ways for techlaw scholarship to structure those lines of inquiry.

71. See, e.g., Anupam Chander, Margot Kaminski & William McGeeveran, *Catalyzing Privacy*, 105 MINN. L. REV. 1733; Margot Kaminski, *A Recent Renaissance in Privacy Law*, 63 COMM'N ACM 24 (2020).

72. See, e.g., Rebecca Hamilton, *Governing the Global Public Square*, 62 HARV. INT'L L.J. 117 (2021); Min Jiang, *Authoritarian Informationalism: China's Approach to Internet Sovereignty*, 30 SAIS REV. INT'L AFF. 71, 71–89 (2010); MARGARET ROBERTS, *CENSORED: DISTRACTION AND DIVERSION INSIDE CHINA'S GREAT FIREWALL* (2019); ZEYNEP TUFECKI, *TWITTER AND TEAR GAS: THE POWER AND FRAGILITY OF NETWORKED PROTEST* (2017).

73. See, e.g., Kim Lyons, *Mark Zuckerberg 'Worried' about China's Influence on Internet Regulation*, THE VERGE (May 18, 2020), <https://www.theverge.com/2020/5/18/21262707/zuckerberg-china-regulation-privacy-facebook>; Nitasha Tiku, *Big Tech; Breaking us up will Only Help China*, WIRED (May 23, 2019), <https://www.wired.com/story/big-tech-breaking-will-only-help-china/>.

IV. THE RULE OF LAW AFTER THE CONTROL REVOLUTION

If the project of techlaw is not to wither into irrelevance as it enters its second quarter century, its core research agenda must concern new organizational forms for legal institutions—organizational forms that are optimized to networked information and communication geographies while reasserting the centrality of public, political authority. As a way of framing that agenda, it is useful to recall Langdon Winner’s important meditation on the possibility of inherently political technologies, which identified nuclear power as a technology that uniquely required authoritarian chains of control.⁷⁴ In her powerful new book, Kate Crawford extends Winner’s point about politics to political economy and to the processes that constitute “artificial intelligence.” She argues that those processes are both inherently authoritarian, because they rely on imposed classification and sorting, and inherently extractive, because of the natural and human resources they demand and consume.⁷⁵ From a lawyer’s perspective, however, Winner’s conclusion about nuclear power was incomplete because authoritarian control processes still might be situated within and subjected to forms of oversight by larger and more democratically accountable institutions. By analogy, it is important to consider not only the modes of control and resource extraction that data-driven, algorithmic processes seem to require in their current implementations, but also whether it might be possible to reconfigure such processes in ways that constrain them to serve democratic, human, and planetary needs.

A. NETWORKED GEOGRAPHIES: MAPPING FLOWS, CONTROL POINTS AND FAILURE MODES

Designing governance institutions capable of subjecting the networked information processes of the control revolution to effective, democratically accountable oversight requires attention to their distinctive geographies—to the patterns of flow they enable, the points of control they offer, and the failure modes they present. The problems that have seemed most unruly when considered from traditional legal-institutional perspectives can help us to surface deeply-rooted assumptions about how a functioning system of legal institutions ought to work, and to define new research agendas that do not take those assumptions as givens.

74. See LANGDON WINNER, *THE WHALE AND THE REACTOR: A SEARCH FOR LIMITS IN AN AGE OF HIGH TECHNOLOGY* 19–39 (1986).

75. See *generally* KATE CRAWFORD, *ATLAS OF AI: POWER, POLITICS, AND THE PLANETARY COSTS OF ARTIFICIAL INTELLIGENCE* (2021).

The first and most basic difficulty that networked information processes have been thought to present for legal institutions is the decentralized, nonhierarchical structure of networks themselves. Networks are not ungovernable, however; they are just governed differently, via the standards that bind participants together and that work—more or less effectively—to prevent defection and alternative network-making.⁷⁶ Standard-based governance mechanisms present distinctive failure modes, some of which relate to hegemonic power, others to defects in mechanisms for inclusion and participation, and others to moral hazard.⁷⁷ Preexisting rule-based legal institutions can amplify the failures. In particular, dominant information platform firms can and do leverage contract, trade secrecy, and intellectual property rights to control access to and uses of their networks.⁷⁸ Legal institutions for the control revolution need not take any of those arrangements for granted, but law- and policymakers must be willing to revise their own assumptions about the primacy and sanctity of contract and property rules designed for an earlier era.

A second set of problems concerns the interdependence of actors and regulatory objects within networked ecosystems. Data-driven predictions derive from and operate on population aggregates, and the scope of effective protection for private information typically depends on the behaviors of relatives, friends, and other members of one's professional and social networks.⁷⁹ Networked information and communication tools have similarly broad affordances and effects.⁸⁰ Legal theories of causation and duty handle such network effects poorly, framing them as externalities and sharply limiting opportunities to impose corrective obligations on those whose conduct creates diffuse external harms. Grants of regulatory jurisdiction designed for a

76. For a useful summary of the different ways that networks reflect and reproduce power, see MANUEL CASTELLS, *COMMUNICATION POWER* 45–46 (2nd ed. 2011).

77. See COHEN, *supra* note 8, at 217–37. See generally THE LAW, ECONOMICS, AND POLITICS OF INTERNATIONAL STANDARDIZATION (Panagiotis Delimatsis ed., 2015); Melissa J. Durkee, *International Lobbying Law*, 127 *Yale L.J.* (2018); Ruth W. Grant & Robert O. Keohane, *Accountability and Abuses of Power in World Politics*, 99 *AM. POLIT. SCI. REV.* 29 (2005); Hans Krause Hansen & Tony Porter, *What Do Numbers Do in Transnational Governance?*, 6 *INT'L POL. SOCIO.* 409 (2012); DAVID SINGH GREWAL, *NETWORK POWER: THE SOCIAL DYNAMICS OF GLOBALIZATION* 25–43, 193–214 (2008).

78. See COHEN, *supra* note 8, at 40–46.

79. See generally Solon Barocas & Karen Levy, *Privacy Dependencies*, 95 *WASH. L. REV.* 555 (2020).

80. See generally Seda Gürses & Joris van Hoboken, *Privacy after the Agile Turn*, in *CAMBRIDGE HANDBOOK OF CONSUMER PRIVACY* 579 (Evan Selinger, Jules Polonetsky & Omer Tene eds., 2018); Karen E.C. Levy, *The User as Network*, 20 *FIRST MONDAY* (2015), <https://doi.org/10.5210/fm.v20i11.6281>.

previous era (and often further constrained by insistently deregulatory approaches to cost-benefit analysis) do not perform much better.⁸¹ Governance institutions for the control revolution require more sophisticated understandings of collective harm and obligation, and of the ways that design interventions can protect both individual and collective values.⁸²

Third and relatedly, as described in Section III.C, digital processes that operate via machine learning detect preexisting patterns in the data upon which they rely and, unless constrained to behave differently, will reproduce those patterns along with whatever biases and systemic injustices they encode. If such processes are not merely to be mechanisms for further entrenching inequality and injustice, law- and policymakers will need to learn to make different uses of what they reveal and must stand ready to reconsider the law's relationships to a wide variety of institutions and practices, many of which are decades and even centuries old.⁸³

A fourth set of problems revolves around the fact that access to networked processes and services is unavoidably mediated in ways designed and controlled by others. Even technically trained experts cannot fully access the details of complex machine learning processes that operate in real time over very large, heterogeneous data sets. The rest of us experience such processes and services (and can hold them accountable) only via interfaces, indicators, and dashboards that communicate selected items of information about how they operate. The traditional criteria relied on by regulators and judges tend not to make sense in such environments—information is always imperfect, choices are always imposed by others, and autonomy is always only partial. Interfaces, indicators, and dashboards also have distinctive failure modes. These range from dark patterns deliberately designed to deceive users to interface conventions designed for their addictive properties to simplifying conventions that reflect incomplete and self-interested perceptions of relevancy and risk.⁸⁴ Governance institutions for the control revolution will

81. See COHEN, *supra* note 8, at 146–54, 173–85.

82. See generally Salomé Viljoen, *A Relational Theory of Data Governance*, 131 YALE L.J. 573 (2021).

83. For three very different approaches to questions involving the legitimate design of predictive algorithmic tools, see Crystal S. Yang & Will Dobbie, *Equal Protection Under Algorithms: A New Statistical and Legal Framework*, 119 MICH. L. REV. 291 (2020); Pauline Kim, *Race-Aware Algorithms: Fairness, Nondiscrimination, and Affirmative Action* (May 2021) (unpublished manuscript) (on file with author); and Deborah Hellman, *Big Data and Compounding Injustice*, 18 J. MORAL PHIL. (forthcoming 2021).

84. On dark patterns, see generally WOODROW HARTZOG, *PRIVACY'S BLUEPRINT: THE BATTLE TO CONTROL THE DESIGN OF NEW TECHNOLOGIES* (2018); Arunesh Mathur, Gunes Acar, Michael J. Friedman, Elena Lucherini, Jonathan Mayer, Marshini Chetty & Arvind Narayanan, *Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites*, 3 PROC. ACM

need to speak the relevant technical and design languages and to open underlying design and optimization processes to appropriately structured forms of public scrutiny.

A final and enormously important cluster of problems involves scale and amplification. Networked digital processes operate and are designed to operate at scale, and their dysfunctions also scale up commensurately. Existing legal institutions do not adequately reckon with scale-based effects or the processes of data-driven intermediation that produce and entrench them. So for example, doctrinal frameworks for common carriage and contributory liability remain tethered to outdated notions of neutrality and fault, but data-driven algorithmic processes chart a middle path, rearranging online interactions in ways that are neither neutral nor intentional but rather driven by instrumental considerations and optimization parameters.⁸⁵ First Amendment doctrine presumes listener rationality and holds that the costs of mistakenly suppressing protected speech outweigh those of mistakenly allowing unprotected speech to spread, but data-driven algorithmic processes optimized for engagement and virality short-circuit the presumptive self-correction mechanisms of hypothesized speech markets.⁸⁶ Theories of privacy oriented toward individual control rights and litigation-centered enforcement cannot constrain data harvesting and processing practices designed to operate on populations.⁸⁷ Governance institutions for the control revolution should be designed in ways that respond to these dynamics.⁸⁸

B. LEGAL NORMATIVITIES: COUNTERING SYSTEMATIC ABUSES OF POWER

One might wonder whether new forms of governance constructed along the lines sketched above still deserve to be called “legal” at all—and why that

HUM-COMPUT. INTERACTIONS 1 (2019). On addictive design, see generally ADAM ALTER, *IRRESISTIBLE: THE RISE OF ADDICTIVE TECHNOLOGY AND THE BUSINESS OF KEEPING US HOOKED* (2017); ZUBOFF, *supra* note 52, at 159–62, 457–74; Tristan Harris, *The Slot Machine in Your Pocket*, DER SPIEGEL ONLINE (July 27, 2016), <http://www.spiegel.de/international/zeitgeist/smartphone-addiction-is-part-of-the-design-a-1104237.html>.

85. See generally Cohen, *supra* note 38.

86. See *id.* at 649–53; Lyriisa Barnett Lidsky, *Nobody’s Fools: The Rational Audience as First Amendment Ideal*, 2010 U. ILL. L. REV. 799 (2010); Wu, *supra* note 39; HENRY FARRELL & BRUCE SCHNEIER, *COMMON-KNOWLEDGE ATTACKS ON DEMOCRACY* (Berkman Klein Ctr. Research Publication No. 2018-7, 2018).

87. See generally Julie E. Cohen, *How (Not) to Write a Privacy Law*, KNIGHT FIRST AMENDMENT INST. (Mar. 23, 2021), <https://knightcolumbia.org/content/how-not-to-write-a-privacy-law>.

88. For preliminary explorations of topics relating to scale-based governance, see *id.*; Paul Ohm, *Regulating at Scale*, 2 GEO. L. TECH. REV. 546 (2018); Rory Van Loo, *The New Gatekeepers: Private Firms as Public Enforcers*, 106 VA. L. REV. 467 (2020).

designation might matter. According to Mireille Hildebrandt, “law,” as we have customarily understood it, is an artifact of print technologies, and especially of the fixity and the temporal rhythms that they impose.⁸⁹ If that is right, then the project of reconstructing the rule of law for the networked digital era is doomed to failure. Understood more broadly, however, “rule of law” language is intended to supply a framework for talking about power, calling it to account, and constraining its systematic abuse. Designing governance institutions for the networked information era requires new thinking about how to translate those broadly framed rule-of-law commitments into mid-level principles capable of being operationalized within networked digital environments.

Legal philosophers probing below the surface of contemporary rule of law discourses have long recognized that the “rule of law” is an essentially contested concept.⁹⁰ Three features of those debates are worth emphasizing here.

First and most important, rule of law discourses are situated in particular places and times, and so they have tended to privilege correspondingly situated institutional solutions. Hildebrandt links law’s decline to the failure of traditional legitimacy criteria such as generality, stability, and reproducibility within digital environments.⁹¹ Those criteria, however, are bound up with the organizational forms within which they have been articulated and reinforced; in particular, they are designed to privilege oversight by courts. If courts and textual fixity cannot contend with new forms of networked power and their endemic failure modes, it becomes important to consider what new organizational forms and accompanying evaluative practices might be devised. Such forms and criteria might bear only passing or partial resemblances to those with which we have been familiar, but there will be learning from other disciplines (such as information security and quality assurance) that might inform their design.⁹²

89. MIREILLE HILDEBRANDT, SMART TECHNOLOGIES AND THE END(S) OF LAW: NOVEL ENTANGLEMENTS OF LAW AND TECHNOLOGY 174–85 (2015).

90. See generally W.B. Gallie, *Essentially Contested Concepts*, 56 PROC. ARISTOTELIAN SOC. 167 (1956).

91. HILDEBRANDT, *supra* note 89, at 174–85; see also *id.* at 133–56 (discussing rule of law components). Relatedly, though with a narrower, Anglo-American focus, Lewis Kornhauser has suggested that law represents an “achievement” of governance that satisfies certain operational and evaluative criteria. See Lewis A. Kornhauser, *Law as an Achievement of Governance* (NYU Sch. L. Pub. L., Working Paper No. 21-04, 2021), <http://dx.doi.org/10.2139/ssrn.3762033>.

92. For preliminary explorations of topics relating to new governance modalities for data-driven algorithmic systems, see John Bowers, Elaine Sedenberg & Jonathan Zittrain, *Platform Accountability through Digital “Poison Cabinets”*, KNIGHT FIRST AMEND. INST. (Apr. 13,

Second, scholars have long recognized that some formulations of the rule-of-law ideal are exceedingly thin and serve as fig leaves for new concentrations of economic, authoritarian, and kleptocratic power. The mostly performative array of institutions that answer to autocrats and dictators tend to track existing, industrial-era presumptions about the form of legal institutions; it is no accident that authoritarian regimes constitute courts and appoint judges even as they withhold the authority that such entities require. (Those who conceived and designed the Facebook Oversight Board to cater to a modern-day autocrat's desire for the trappings of the rule of law could have learned a thing or two from this history.) Judged according to the traditional legitimacy criteria, however, such regimes implement the rule of law in name only. Responses to such efforts emphasize the importance of higher-level evaluative criteria such as, for example, tempering arbitrary power.⁹³ Power is resourceful, the project of tempering it is ongoing, and experiments in legal-institutional design will be more or less successful in that regard. Returning iteratively to higher-level rule of law criteria will be important in assessing the resilience and durability of new legal-institutional governance formations.

Finally, rule of law discourses can appear to sanction results that, while nonarbitrary, are nonetheless deeply unjust. Our notions of merit and fault as essentially individualized attributes have produced widespread acceptance of legal-institutional practices that satisfy the traditional criteria of regularity and publicity—and so, not coincidentally, may be consistent with fig-leaf accounts of what the rule of law requires—but that are designed to *further* and *widen* systematically unequal resource distribution. I have in mind here a wide and varied set of practices including, for example, land use and public education regulations that reinforce patterns of economic inequality; rules for conducting elections that, while formally neutral, produce systematic disparities in voter turnout and political representation; and rules regarding the imposition and enforcement of legal financial obligations that impose crushing burdens on the poor.⁹⁴ In contrast, Paul Gowder's exploration of the rule of law foregrounds an equality criterion and demands that rule-of-law institutions work to

2021), <https://knightcolumbia.org/content/platform-accountability-through-digital-poison-cabinets>; douek, *Second Wave Content Moderation*, *supra* note 68; Goldman, *supra* note 65; Paul Ohm, *Sensitive Information*, 88 S. CALIF. L. REV. 1125 (2015); Lauren E. Willis, *Performance-Based Consumer Regulation*, 82 U. CHI. L. REV. 1309 (2015).

93. See, e.g., Martin Krygier, *The Rule of Law: Pasts, Presents, and Two Possible Futures*, 12 ANN. REV. L. & SOC. SCI. 199 (2016).

94. On the term "legal financial obligations," see Monica Llorente, *Criminalizing Poverty through Fines, Fees, and Costs*, AM. BAR ASS'N (Oct. 3, 2016), <https://www.americanbar.org/groups/litigation/committees/childrens-rights/articles/2016/criminalizing-poverty-fines-fees-costs/>.

counteract strategies for hoarding perks and privileges.⁹⁵ A rule-of-law framework for a post- and decolonial era might—and, I would argue, should—give much greater weight to such considerations. A rule of law framework for the networked information era should include mid-level principles for operationalizing an equality criterion within networked digital environments. In particular, it should recognize that those with greater access to knowledge and processing power will always be able to take advantage of information gaps, and that considerations of systemic, distributive, and intergenerational justice may require leveling interventions.⁹⁶

V. CONCLUSION: WWJD?

The legal academy and the legal profession now confront a generational challenge. It is useful to begin simply by recognizing as much. In the context of this symposium, it is also both fitting and instructive to return, once again, to “Lex Informatica” and to Joel Reidenberg. What would Joel do?

That question is easy to answer: Look past overly reductive models and pat solutions. Center legal institutions as necessary sites of innovation. Consult technologists, but don’t conflate their particular expertise with wisdom about how to run a just, inclusive, and democratically-accountable society.⁹⁷ Consult industry, but don’t confuse its self-interested, ideologically overdetermined positionings about “progress” and “innovation” with the demands of human flourishing more broadly understood.⁹⁸ Design processes that prioritize *public* accountability.⁹⁹ Especially, prioritize accountability to communities that have borne the brunt of legally- and technologically-facilitated abuses.¹⁰⁰ Bring

95. See generally PAUL GOWDER, *THE RULE OF LAW IN THE REAL WORLD* (2016).

96. On the consequences of differential access to information and processing power, see generally MARK ANDREJEVIC, *INFOGLUT: HOW TOO MUCH INFORMATION IS CHANGING THE WAY WE THINK AND KNOW* (2013); Hellman, *supra* note 83. For preliminary explorations of topics relating to leveling interventions in systems design, see Julie E. Cohen, *Turning Privacy Inside Out*, 20 *THEOR. INQ. L.* 1 (2019); Mireille Hildebrandt, *Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning*, 20 *THEOR. INQ. L.* 83 (2019); Paul Ohm & Jonathan Frankle, *Desirable Inefficiency*, 70 *FLA. L. REV.* 777 (2018); Paul Ohm, *Forthright Code*, 56 *HOUS. L. REV.* 471 (2018).

97. See Joshua A. Kroll, Joanna Huey, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson & Harlan Yu, *Accountable Algorithms*, 165 *U. PA. L. REV.* 633, 695–99 (2017).

98. See Joel R. Reidenberg & Thomas H. Davenport, *Should the U.S. Adopt European-Style Data-Privacy Protections?*, *WALL. ST. J.* (Mar. 18, 2013), <https://www.wsj.com/articles/SB10001424127887324338604578328393797127094>.

99. See Kroll et al., *supra* note 97, at 702–05.

100. See *id.* at 678–95; N. Cameron Russell, Joel R. Reidenberg, Elizabeth Martin & Thomas B. Norton, *Transparency and the Marketplace for Student Data*, 22 *VA. J.L. & TECH.* 107 (2018).

everyone to the table, and treat everyone with generosity and respect.¹⁰¹ Above all, remember that law is a means to an end and that denial, defeatism, arrogance, and entitlement undermine that end utterly.

101. See Steve Bellovin, *In Memoriam: Joel Reidenberg*, SMBLOG (Apr. 22, 2020), <https://www.cs.columbia.edu/~smb/blog/2020-04/2020-04-22.html>; Omer Tene, *A Farewell to Joel Reidenberg: Mentor, Scholar, Mensch*, IAPP.ORG (Apr. 23, 2020), <https://iapp.org/news/a/a-farewell-to-joel-reidenberg-mentor-scholar-mensch/>.

RACIAL SEGREGATION AND THE DATA-DRIVEN SOCIETY: HOW OUR FAILURE TO RECKON WITH ROOT CAUSES PERPETUATES SEPARATE AND UNEQUAL REALITIES

Rashida Richardson[†]

ABSTRACT

This Article asserts that in the United States, racial segregation has played and continues to play a central evolutionary role in the inequalities reproduced and amplified by modern data-driven technologies and applications. Racial segregation distorts and constrains the development and implementation of data-driven technologies, conceptualization of algorithmic bias problems, and assessments of interventions or solutions deemed appropriate and worth pursuing. There are three main benefits of critical analysis of racial segregation. First, it can deepen our understanding of algorithmic bias. Second, it can improve evaluations of data-driven technologies for social and racial equity concerns. And, third, it can broaden our imaginations about meaningful redress of technology-mediated harms and injustices. This Article starts with a review of the foundational aspects of the evolution of racial segregation, and the social, political, and epistemic implications of racial segregation for the demographic group that dominates the technology sector—White Americans. This Article then explores how racial segregation affects algorithmic design, analysis, and outcomes. In conclusion, this Article analyzes prevailing approaches to evaluating and mitigating algorithmic bias, demonstrates the insufficiencies of those approaches, and proposes a transformative justice framework to adequately examine and redress algorithmic bias and to improve the development of data-driven technologies and applications.

DOI: <https://doi.org/10.15779/Z38PN8XG3V>

© 2021 Rashida Richardson.

[†] Rashida Richardson is an Assistant Professor of Law and Political Science at Northeastern University. The author appreciates helpful feedback in developing the ideas and analysis presented in this Article or on early drafts from Carl Slater Jr., Stephon Richardson, Kathrine D. Stapleton, Donald Richardson, Wei Tchou, Ángel Díaz, Varoon Mathur, Maurice R. Dyson, Joy Lisi Rankin, B. Nicole Triplett, and Jumana Musa.

TABLE OF CONTENTS

I.	INTRODUCTION	1052
II.	RACIAL SEGREGATION IN THE UNITED STATES AND ITS IMPLICATIONS.....	1059
A.	A BRIEF OVERVIEW OF RACIAL SEGREGATION IN THE UNITED STATES.....	1059
B.	A REVIEW OF THE SOCIAL, POLITICAL, AND EPISTEMIC IMPLICATIONS OF RACIAL SEGREGATION.....	1064
III.	HOW RACIAL SEGREGATION SHAPES DATA-DRIVEN TECHNOLOGIES AND ALGORITHMIC BIAS.....	1070
A.	RACIAL SEGREGATION AND ALGORITHMIC DESIGN.....	1070
1.	<i>Training Data</i>	1071
2.	<i>Problem Formulation</i>	1074
B.	RACIAL SEGREGATION AND ALGORITHMIC EVALUATION.....	1077
IV.	CONCLUSION.....	1085

The interlocking workings of human worth, race, and space demonstrate the ways the uninhabitable still holds currency in the present and continues to organize contemporary geographic arrangements. The colonial enactment of geographic knowledge mapped “a normal way of life” through measuring different degrees of humanness and attaching different versions of the human to different places.

—Katherine McKittrick, *Plantation Futures*

I. INTRODUCTION

The United States is one of the most racially diverse nations, leading in the development, use, and evaluation of data-driven technologies and methods. Yet, the diversity of the United States is not reflected in its technology sector. As shown by several years of annual diversity reports, large technology companies like Amazon and Facebook have made little progress in hiring and retaining underrepresented racial minorities.¹ When it comes to leadership

1. Kate Rooney & Yasmin Khorram, *Tech Companies Say They Value Diversity but Reports Show Little Change in Last Six Years*, CNBC (June 12, 2020, 11:27 AM), <https://www.cnbc.com/2020/06/12/six-years-into-diversity-reports-big-tech-has-made-little-progress.html> (highlighting recent diversity reports of technology companies and a dismal improvement they made in recruiting and retaining underrepresented demographics); U.S.

positions, the picture is even bleaker.² Similarly, academic institutions are failing to engage and retain racially diverse faculty and students in science, technology, engineering, and mathematics (STEM) fields. Although Black, Indigenous, and other people of color (BIPOC) make up almost forty percent of the U.S. population,³ those racial minorities make up only twenty-two percent of STEM bachelors, nine percent of STEM doctorates, and ten percent of STEM faculty at four-year institutions.⁴ This lack of racial diversity extends to technology policymaking, particularly to government and civil society, which both influence the use, evaluation, and potential oversight of data-driven technologies. To that point, recent studies have found a lack of racial diversity amongst congressional staff and within technology policy careers and institutions, for instance, think tanks and public interest nonprofit organizations.⁵ This lack of diversity coupled with predominately White and male leadership has led some to criticize the technology sector as having a “White guy problem,” rather than a “minority deficiency,” which tends to emphasize meritocracy concerns over systemic or institutional bias issues.⁶

EQUAL EMP. OPPORTUNITY COMM’N, DIVERSITY IN HIGH TECH (2016), <https://www.eeoc.gov/special-report/diversity-high-tech> (detailing demographic trends in the high tech industry and reasons for its lack of diversity).

2. See Rooney & Khorram, *supra*, note 1; U.S. EQUAL EMP. OPPORTUNITY COMM’N, *supra*, note 1.

3. U.S. CENSUS, QUICK FACTS, UNITED STATES (2019), <https://www.census.gov/quickfacts/fact/table/US/PST045219>.

4. NAT’L SCI. FOUND., WOMEN, MINORITIES, AND PERSONS WITH DISABILITIES IN SCIENCE AND ENGINEERING (2019), <https://ncses.nsf.gov/pubs/nsf19304/digest/field-of-degree-minorities>.

5. See PUBLIC KNOWLEDGE, DIVERSITY IN EARLY-CAREER TECH POLICY ROLES: CHALLENGES AND OPPORTUNITIES 7 (2021), https://www.publicknowledge.org/wp-content/uploads/2021/01/Diversity-in-Early-Career-Tech-Policy-Roles_Public-Knowledge.pdf (citing studies finding that people of color make up only eleven percent of top Senate personal offices and only seventeen percent of senior leadership in nonprofit organizations); LASHONDA BRENSON, JOINT CTR. FOR POL. & ECON. STUD., RACIAL DIVERSITY AMONG TOP STAFF IN SENATE PERSONAL OFFICES 2-3 (2020), https://jointcenter.org/wp-content/uploads/2020/08/2020-Senate-Report-Draft__08-21-20-5AM.pdf (people of color account for only nine percent of U.S. Senators and eleven percent of top staff in the personal offices of U.S. Senators); ELSIE L. SCOTT, KARRA W. MCCRAY, DONALD BELL & SPENCER OVERTON, JOINT CTR. FOR POL. & ECON. STUD., RACIAL DIVERSITY AMONG TOP U.S. HOUSE STAFF 5 (2018), <https://jointcenter.org/wp-content/uploads/2019/11/Racial-Diversity-Among-Top-US-House-Staff-9-11-18-245pm-1.pdf> (stating the U.S. House of Representatives staffers of color account for only 13.7% of top staff positions).

6. See Karen Hao, *AI’s White Guy Problem Isn’t Going Away*, MIT TECH. REV. (Apr. 17, 2019), <https://www.technologyreview.com/2019/04/17/136072/ais-White-guy-problem-isnt-going-away> (highlighting research on the consequences of the lack of diversity and having a predominately White, male leadership in the technology sector); Michael Schaus, *Yahoo*

This problem of racial homogeneity in the technology sector is not new. Yet, scholarship and public discourse regarding racial segregation's impact on technology development and algorithmic bias remains sparse.⁷ Segregation refers to a systematic spatial separation and social exclusion of groups, and in the United States, segregation has primarily and consistently occurred on the basis of race and ethnicity. For decades, American scholars have referred to two types of racial segregation in the United States: *de jure* and *de facto*.⁸

De jure segregation is a legally mandated separation and social regulation of races. This form of segregation was imposed through explicitly racially discriminatory laws and regulations, such as slave codes, Federal Indian policy, Black Codes, and Jim Crow laws. Though *de jure* segregation was eventually outlawed through a series of U.S. Supreme Court cases and federal civil rights laws, remnants of these laws and their intended consequences are perpetuated today through political beliefs and practices like “local control.”⁹ Moreover, those remnants and consequences persist because some racially neutral laws

Liberals: Apple's "Old White Guy" Problem, TOWNHALL FIN. (Oct. 25, 2013, 1:20 AM), <https://finance.townhall.com/columnists/michaelschaus/2013/10/25/yahoo-liberals-apples-old-white-guy-problem-n1731892> (criticizing media for framing technology sector issues as an “old White guy problem” and insisting that it is rather a “minority deficiency” issue); Michelle Toh, *Ellen Pao: Meritocracy in Tech is a Myth*, CNN (Apr. 21, 2021, 8:35 AM), <https://www.cnn.com/2021/04/21/tech/ellen-pao-anti-asian-hate-intl-hnk/index.html> (arguing that “[t]he ‘biggest myth’ about working in the tech[nology] industry is that it is based on meritocracy” based on the experience of the former Reddit CEO Ellen Pao); Mar Hicks, *A Feature, Not a Bug*, 5 TECH. STORIES (Dec. 4, 2017), <http://www.technologystories.org/a-feature-not-a-bug> (arguing that notions of meritocracy in Silicon Valley disguise systemic problems like racism and sexism).

7. See, e.g., Margaret Hu, *Algorithmic Jim Crow*, 86 FORDHAM L. REV. 633, 633–71 (2017) (evaluating how algorithmic tools—such as current immigration and security-related vetting protocols—replicate policies and practices that comprised the Jim Crow regime).

8. E.g., Katie R. Eyer, *Ideological Drift and the Forgotten History of Intent*, 51 HARV. C.R.-C.L. L. REV. 1, 25 n.131 (2016); Robert L. Carter, *De Facto School Segregation: An Examination of the Legal and Constitutional Questions Presented*, 16 CASE WESTERN RES. L. REV. 502, 503 (1965). U.S. courts occasionally rely on this distinction as well; see, e.g., *Parents Involved in Cmty. Sch. v. Seattle Sch. Dist. No. 1*, 551 U.S. 701, 794 (2007) (Kennedy, J. dissenting).

9. Local control refers to the delegation of authority in governance, management, and decision-making to local governments or local governing bodies. This position also assumes that state and federal authorities should not interfere with or attempt to check local government policies or practices. It is commonly used in the education sector in reference to public schools’ governance and management. See generally SUSAN E. EATON & ELIZABETH CRUTCHER, *THE HARVARD PROJECT ON SCHOOL DESEGREGATION, SLIPPING TOWARDS SEGREGATION: LOCAL CONTROL AND ERODING DESEGREGATION IN MONTGOMERY COUNTY, MARYLAND* (1994).

and policies, like taxation, are designed with only White Americans in mind and ignore “the reality of societal differences based on race.”¹⁰

De facto segregation, on the other hand, is the social exclusion and regulation of races without legal mandates. Instead, this form of segregation exists through social customs, voluntary practices, private discriminations, and other prejudicial practices or behaviors.¹¹ Though some scholars and practitioners contend that the distinction between de jure and de facto segregation is fallacious,¹² there is general consensus that racial segregation remains one of the most persistent and pervasive features of American society.¹³ In fact, “hypersegregation” has become a prominent term used to describe the modern-day, multidimensional nature of racial segregation.¹⁴ So, although the goal of racial segregation was to create spatial separation and social exclusion, its immediate and long-term consequences have been the entrenchment of racial inequality across all facets of society.¹⁵

When the impact of racial segregation is ignored, issues of racial inequality appear as naturally occurring phenomena, rather than byproducts of specific

10. DOROTHY A. BROWN, *THE WHITENESS OF WEALTH: HOW THE TAX SYSTEM IMPOVERISHES BLACK AMERICANS AND HOW WE CAN FIX IT* 9–24 (2021) (describing how various tax laws and policies contribute to racial disparities in wealth accumulation, patterns of economic distribution, and socioeconomic status).

11. *See Parents Involved in Cmty. Sch.*, 551 U.S. at 794 (Kennedy, J. dissenting) (detailing the differences between *de jure* and *de facto* segregation).

12. *See generally* RICHARD ROTHSTEIN, *THE COLOR OF LAW: A FORGOTTEN HISTORY OF HOW OUR GOVERNMENT SEGREGATED AMERICA* (2017) (describing how many historical instances of de facto segregation were, in fact, the result of public policies and political practices); DERRICK BELL, *SILENT COVENANTS: BROWN V. BOARD OF EDUCATION AND THE UNFULFILLED HOPES FOR RACIAL REFORM* (2004) (highlighting that racial policies, like racial segregation, are made through silent covenants—unspoken convergences of interest and involuntary sacrifices of rights—that ensure that policies conform to priorities set by policymakers).

13. *E.g.*, Paul A. Jargowsky, *The Persistence of Segregation in the 21st Century*, 36 MINN. J.L. & INEQ. 207 (2018); Greg Rosalsky, *What a 1968 Report Tells Us About the Persistence of Racial Inequality*, NPR (June 9, 2020, 6:30 AM), <https://www.npr.org/sections/money/2020/06/09/872402262/what-a-1968-report-tells-us-about-the-persistence-of-racial-inequality>. *See generally* *The Dream Revisited*, NYU FURMAN CTR., <https://furmancenter.org/research/iri/discussions> (last visited July 11, 2021) (series discussing the cross-sections of segregation and policy areas such as finance, affordable housing, and poverty).

14. “Hypersegregation” is a term that describes segregation of a race or ethnic group in multiple ways or across a range of measures. The concept was introduced and advanced by sociologists Douglas S. Massey and Nancy A. Denton in their research of the unique form of segregation that concentrated large proportions of Black Americans in poor urban neighborhoods. *See generally* Douglas S. Massey & Nancy A. Denton, *Hypersegregation in the U.S. Metropolitan Areas: Black and Hispanic Segregation Along Five Dimensions*, 26 DEMOGRAPHY 373 (1989).

15. *See* NYU FURMAN CTR., *supra* note 13.

policies, practices, social norms, and behaviors. Ignoring the impact of racial segregation also means that historical contingencies remain erroneously timebound to actors and institutions of the past, leaving unexamined and unchanged individual and collective practices that perpetuate or even exacerbate racial inequalities. This is perhaps why racial inequalities within the technology sector and racially biased outcomes of algorithmic technologies persist. It is impossible to improve racial diversity in STEM fields without critically examining the persistence of racial segregation in American public schools, particularly in regards to academic tracking, school funding, and school discipline.¹⁶ It is also impossible to improve recruitment and retention of underrepresented racial minorities in the technology sector without critically examining how economic, occupational, and residential segregation may deter minority candidates from applying for jobs while undermining recruitment efforts within the sector.¹⁷ And it is further impossible to mitigate algorithmic bias without critically examining how training datasets may be systemically biased by racial segregation, or why primarily technical solutions, such as audits, cannot meaningfully redress problems caused by structural inequality.

But focusing on racial segregation can help mitigate those issues in two ways. First, examining racial segregation may explain uneven progress—and, in some cases, regression—in addressing this form of structural inequality and its manifestations in particular sectors like the technology industry. And second, racial segregation reveals the unstated, and often unconscious, motivations, interests, and practices that make racial segregation and its attendant consequences so resilient in American society.

This Article contends that racial segregation has played a central evolutionary role in the reproduction and amplification of racial stratification in data-driven technologies and applications. Racial segregation also constrains conceptualization of algorithmic bias problems and relevant interventions. This Article does not merely discuss how racial segregation reflects historical and contemporary patterns of private and public racial discrimination in various social contexts like education, housing, employment, and public goods.

16. See, e.g., Joy Lisi Rankin, *Whitewashing Tech: Why the Erasure of the Past Matter Today*, MEDIUM (Oct. 1, 2020), <https://medium.com/@AINowInstitute/whitewashing-tech-why-the-erasures-of-the-past-matter-today-166d0d5e2789> (“[A]s computing and related STEM fields have become more prestigious, women, BIPOC, and others who’ve faced discrimination have been actively pushed out.”).

17. See, e.g., Quincy Brown, Tyrone Grandison, Jamika D. Burge, Odest Chadwicke Jenkins & Tawanna Dillahunt, *Amplifying Resources for Inclusiveness in Computing: Reflections on Black in Computing*, COMPUTING RES. ASS’N BULL. (Mar. 31, 2021), <https://cra.org/amplifying-resources-for-inclusiveness-in-computing-reflections-on-black-in-computing>.

These patterns and the societal inequality they produce are well-documented¹⁸ and will be further explored in my future scholarship. Rather, this Article focuses on how the compounded effects of social exclusion and spatial isolation produced by racial segregation affect technology development and algorithmic bias. Those effects include concentrated wealth, distressed communities, and cognitive oversights.

Part II provides a brief historical overview of racial segregation, the more obvious racial inequities it has produced, and its less obvious social, political, and epistemic implications for White Americans. Since White Americans dominate the technology sector¹⁹ and, as most research suggests, primarily benefit from racial segregation,²⁰ it is important to evaluate White Americans' relationship to the problem before examining how racial segregation affects algorithmic designs, analyses, and outcomes.

Part III explores how racial segregation impacts data-driven technologies and algorithmic bias in two aspects. First, Section III.A. explores how racial segregation and the inequality it breeds influences algorithmic design. Here, the Article highlights specific approaches and presumptions in algorithmic design and analysis that tend to replicate and maintain racial inequalities produced by racial segregation. Second, Section III.B. explores how racial segregation, or the failure to account for it, influences the evaluation of some data-driven technologies. This Part of the Article highlights how socially

18. See generally ROTHSTEIN, *supra* note 12; JESSICA TROUNSTINE, *SEGREGATION BY DESIGN: LOCAL POLITICS AND INEQUALITY IN AMERICAN CITIES* (2018) (analyzing more than 100 years of data from American cities and how such cities created racial segregation through local governance); Jennifer Roback, *Southern Labor Law in the Jim Crow Era: Exploitive or Competitive?*, U. CHI. L. REV. 1161 (1984) (exploring the economic effects of Southern labor laws during the Jim Crow era).

19. E.g., Sinduja Rangarajan, *Here's the Clearest Picture of Silicon Valley's Diversity Yet: It's Bad. But Some Companies are Doing Less Bad*, REVEAL (June 25, 2018), https://revealnews.org/article/heres-the-clearest-picture-of-silicon-valleys-diversity-yet/?utm_source=Reveal&utm_medium=social_media&utm_campaign=twitter (finding that White men make up the majority of the Silicon Valley's professional workforce, especially at the managerial and executive levels); Joe Davidson, *Mostly White Male Tech Sector Needs Government Help on Diversity*, WASH. POST (Dec. 4, 2017), <https://www.washingtonpost.com/news/powerpost/wp/2017/12/04/tech-sector-needs-uncle-sams-help-on-diversity/> ("Not only is it made up overwhelmingly of White men, but the percentage of tech workers who are black decreased in recent years, while the portion of women in the industry was stagnant and the level of Hispanic workers was nearly flat.").

20. See generally IRA KATZNELSON, *WHEN AFFIRMATIVE ACTION WAS WHITE* (2006); TROUNSTINE, *supra* note 18; EDUARDO BONILLA-SILVA, *RACISM WITHOUT RACISTS: COLOR-BLIND RACISM AND THE PERSISTENCE OF RACIAL INEQUALITY IN AMERICA* (2014) (challenging how White Americans use color-blind racism to frame racial affairs); Daria Roithmayr, *Racial Cartels*, 16 MICH. J. RACE & L. 45 (2010) (noting how historically all-White groups, such as homeowners' associations, have benefitted from such racial exclusions).

contested data-driven technologies are considered, by some, to be fair and permissible. Part III is not intended to comprehensively evaluate the various ways racial segregation influences algorithmic design and bias. Instead, Part III helps illustrate how critical analysis of racial segregation can deepen our understanding of algorithmic bias, improve evaluations of data-driven technologies for social and racial equity concerns, and broaden our imaginations about what meaningful redress of algorithmic bias and racial segregation should include.

The Article concludes in Part IV with an analysis of how prevailing approaches to evaluating and mitigating algorithmic bias are insufficient, why a transformative justice framework is necessary to adequately examine and redress algorithmic bias, and how to improve the development of data-driven technologies and applications.

Though data-driven technologies are developed and used globally, this Article explicitly focuses on the United States because analysis of structural and racial inequality requires specificity. This specificity means examining laws, customs, social practices, and other societal features that are often constrained by or limited to jurisdictional boundaries. However, the thesis and the analysis provided in this Article can have value in other countries and contexts to interrogate local practices or forms of segregation, structural inequality, and systemic oppression (e.g., dispossession, colonization, or caste). Ignoring or masking the origins and contemporary forms of social stratification is not unique to the United States,²¹ but to develop and hone inclusive and accurate global or multi-jurisdictional analysis of algorithmic bias and other technology-related issues, it is essential to understand the nature of the problem locally.

21. *Compare* U.K. COMM'N ON RACE & ETHNIC DISPARITIES, COMMISSION ON RACE AND ETHNIC DISPARITIES: THE REPORT (2021), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/974507/20210331_-_CRED_Report_-_FINAL_-_Web_Accessible.pdf (downplaying the role and impact of racial and ethnic discrimination and institutional racism in modern-day Britain), *with* DOMINIQUE DAY, AHMED REID, SABELO GUMEDZE, MICHAL BALCERZAK, RICARDO A. SUNGAI III & WORKING GROUP OF EXPERTS ON PEOPLE OF AFRICAN DESCENT OF THE SPECIAL PROCEDURES OF THE UNITED NATIONS HUMAN RIGHTS COUNCIL, UN Experts Condemn UK Commission on Race and Ethnic Disparities Report (2021), <https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=27004> (rejecting the findings included in the aforementioned U.K. report because it ignores the pervasive role that race plays in society).

II. RACIAL SEGREGATION IN THE UNITED STATES AND ITS IMPLICATIONS

This Part starts with a high-level overview of how racial segregation was enacted over time in the United States to contextualize the extreme social and spatial isolation racial segregation has produced. With this context, Section II.B then explores the less obvious social, political, and epistemic implications of racial segregation, particularly, for White Americans.

A. A BRIEF OVERVIEW OF RACIAL SEGREGATION IN THE UNITED STATES

Throughout U.S. history, racial segregation has been a primary mechanism for maintaining racial hierarchy. This social and racial paradigm of hierarchy, separation, and exclusion has been primarily maintained through the laws, policies, and actions of private and public institutions and individuals. However, the approaches and means of racial segregation have evolved over time to accommodate legal and societal changes.²²

During the colonial and antebellum periods, racial groups were in closer proximity than in later centuries due to the cruel, yet legal institution of chattel slavery. Instead of being spatially segregated, Black Americans in particular were heavily restricted in their movements, actions, liberties, and overall existence through the enforcement of slave codes (laws regulating slavery and enslaved people in all colonies) as well as Black Codes (laws regulating the activities and behavior of Black Americans during the antebellum era and Reconstruction).²³ Legal restrictions imposed by these laws segregated some public spaces, significantly controlled or limited interracial interactions, and subjugated Black Americans to exploitive work arrangements and conditions even in states where slavery was illegal. The restrictions also had a serious effect on wealth generation by prohibiting Black Americans from settling in new states or territories that were dispossessed from Indigenous Americans.²⁴

22. See NYU FURMAN CTR., *supra* note 13.

23. See generally A. LEON HIGGINBOTHAM JR., *IN THE MATTER OF COLOR: RACE AND THE AMERICAN LEGAL PROCESS: THE COLONIAL PERIOD* (1978) (detailing colonial laws, policies, practices, and cases regarding the regulation of Black and Indigenous Americans); SIMONE BROWN, *DARK MATTERS: ON SURVEILLANCE OF BLACKNESS* (2015) (describing how racialized surveillance is based on and evolved from policing practices developed during the antebellum and Reconstruction periods); ERIC FONER, *THE SECOND FOUNDING: HOW THE CIVIL WAR AND RECONSTRUCTION REMADE THE CONSTITUTION* (2019) (detailing the historical and political context surrounding the development of the Reconstruction constitutional amendments).

24. This exclusion of non-White groups is notable because the Homestead Acts, a series of laws granting government land to citizens by application, were the most extensive

After Emancipation and Reconstruction, racial segregation increased dramatically due to the establishment of the Jim Crow system. The Jim Crow system was a series of state and municipal laws, ordinances, informal policies, extrajudicial practices, and even social customs that required racial separation and discrimination.²⁵ The system existed predominately but not exclusively in the American South.²⁶ This system legalized racial segregation across most aspects of society to serve as “public symbols and constant reminders of [Black Americans’] inferior position.”²⁷ Though the immediate purpose of the Jim Crow system was to enforce physical separation based on race, the system also served to limit or exclude Black Americans from social, political, economic, and legal participation in society. American historian C. Vann Woodward illustrates this point:

The [segregation] code lent the sanction of law to a racial ostracism that extended to churches and schools, to housing and jobs, to eating and drinking. Whether by law or custom, that ostracism extended to virtually all forms of public transportation, to sports and recreations, to hospitals, orphanages, prisons, and asylums, and ultimately to funeral homes, morgues, and cemeteries.²⁸

During the same period, the United States experienced significant industrial growth that spurred massive demographic shifts and urban population growth from domestic migration and immigration.²⁹ Racial segregation in these urban and industrialized areas was less overt, except for exclusionary zoning ordinances, which prohibited or otherwise significantly restricted the

redistributive government policy and one of the greatest wealth generating entitlement programs in U.S. history. See Keri Leigh Merritt, *Land and the Roots of African-American Poverty*, AEON (Mar. 11, 2016), <https://aeon.co/ideas/land-and-the-roots-of-african-american-poverty>. See generally ROXANNE DUNBAR-ORTIZ, AN INDIGENOUS PEOPLES’ HISTORY OF THE UNITED STATES (2015) (showing how U.S. policies have been designed to seize Indigenous peoples’ territories and how those policies have displaced such original inhabitants).

25. Hu, *supra* note 7, at 633.

26. See C.R. & RESTORATIVE JUST. PROJECT, *The CRRJ Burnham-Nobles Archive*, [https://crrj.org/reading-room/\(holding over 1150 documents and cases on racially motivated homicides in the United States South\)](https://crrj.org/reading-room/(holding+over+1150+documents+and+cases+on+racially+motivated+homicides+in+the+United+States+South)); Andrew W. Kahrl, *The North’s Jim Crow*, N.Y. TIMES (May 27, 2018), <https://www.nytimes.com/2018/05/27/opinion/jim-crow-north.html> (describing Jim Crow practices and laws in northern states like Connecticut and New Jersey). See generally LYNN M. HUDSON, WEST OF JIM CROW (2020) (detailing Jim Crow laws, practices, and customs in California).

27. C. VANN WOODWARD, THE STRANGE CAREER OF JIM CROW 7 (commemorative ed. 2002).

28. *Id.*

29. See generally TROUNSTINE, *supra* note 18.

presences of racial minorities.³⁰ Yet, state and local governments encouraged and incentivized racial segregation and homogeneity in two subtle ways. First, they increased spending on public goods, services, and infrastructure, but these investments were made almost exclusively in White neighborhoods.³¹ The public goods, services, and infrastructure spending included basic necessities from sewer and streetlights to public schools and fire departments.³² Second, policymakers and social welfare workers sought to assimilate and “Americanize” foreign-born populations, which were primarily European due to racially exclusionary immigration policies, through expanding their access to these improved public goods, services, and infrastructure.³³ But such increased spending and expanded services were often dependent on and linked to community homogeneity, enabling and masking the systemic denial of public goods and services based on race.³⁴ As a result, non-White neighborhoods became neglected and overcrowded, and they often lacked access to basic public services, like clean water and sewer access, even though these neighborhoods paid their fair share of taxes.³⁵ The squalid conditions of non-White neighborhoods were used to reinforce notions of racial inferiority that in turn helped justify additional discriminatory policies and practices: ignoring crime in Black areas, zoning Black neighborhoods in floodplains,

30. See Christopher Silver, *The Racial Origins of Zoning in American Cities*, in *URBAN PLANNING AND THE AFRICAN AMERICAN COMMUNITY: IN THE SHADOWS* (June Manning Thomas & Marsha Ritzdorf eds., 2008) (noting how race-based urban planning have negatively impacted Black communities); Rolf Pendall, *Local Land Use Regulation and the Chain of Exclusion*, 66 J. AM. PLAN. ASS'N 125, 125–42 (2007).

31. See TROUNSTINE, *supra* note 18, at 98–118 (detailing how urban cities’ infrastructure and public goods were unevenly developed to covertly advance racial segregation interests). See generally CRAIG M. BROWN & CHARLES N. HALABY, *BOSSSES, REFORM, AND THE SOCIOECONOMIC BASES OF URBAN EXPENDITURE, 1890–1940*, in *THE POLITICS OF URBAN FISCAL POLICY* 596–611 (Terrance J. McDonald & Sally K. Ward eds., 1984) (describing how political bosses drove municipal spending to advance local White interests); CYBELLE FOX, *THREE WORLDS OF RELIEF: RACE, IMMIGRATION, AND THE AMERICAN WELFARE STATE FROM THE PROGRESSIVE ERA TO THE NEW DEAL* 1–72 (2012) (finding that despite rampant nativism, European immigrants received generous access to social welfare programs whereas Blacks and Mexicans were not provided access to these programs and benefits in addition to facing more punitive outcomes like aggressive policing and deportation).

32. See sources cited *supra* note 31.

33. See sources cited *supra* note 31.

34. See, e.g., TROUNSTINE, *supra* note 18, at 98–118; BROWN & HALABY, *supra* note 31. See generally Claudia Goldin & Lawrence F. Katz, *Human Capital and Social Capital: The Rise of Secondary Schooling in America*, 29 J. INTERDISC. HIST. 683 (1999).

35. TROUNSTINE, *supra* note 18, at 100–18.

expulsive zoning,³⁶ and urban renewal.³⁷ The totalizing effect of these policies and practices not only made their racialized consequences resilient through time and demographic shifts, but this effect made the full democratic participation or assimilation of racial minorities insuperable.

Federal public policy also contributed to racial segregation nationally in several ways. First, President Franklin D. Roosevelt's administration created many economic and social programs and made many reforms that included provisions or carve-outs that effectively excluded racial minorities and enabled racial segregation. For example, the National Labor Relations Act of 1935 instituted various rights and protections for private sector employees, and the Fair Labor Standards Act of 1938 boosted wages and improved working conditions. Yet, both acts excluded domestic and agricultural workers, the majority of whom were Black Americans, and gave way to several forms of occupational and economic segregation.³⁸ Another way federal policy tacitly sanctioned racial segregation was through relegating implementation of federal programs like the Servicemen's Readjustment Act of 1944 ("G.I. Bill") to state and local governments, many of which upheld the Jim Crow system or other racially discriminatory policies and practices.³⁹ Finally, several of the newly created federal executive agencies and institutions, like the Federal Housing Administration and the Home Owners' Loan Corporation, enacted policies that promoted racial segregation and incentivized racially discriminatory practices and customs (for example, redlining).⁴⁰ Yet, as legal and social challenges to racial segregation mounted and became increasingly successful during the Civil Rights Era, the policies and practices of federal, state, and local governments became more covert. For example, federal authorities used urban renewal and highway development projects to decimate BIPOC communities

36. Expulsive zoning refers to the placement of negative industrial or commercial uses, like a waste incineration plant, in non-White neighborhoods to encourage the displacement of existing residents. See YALE RABIN, *EXPULSIVE ZONING: THE INEQUITABLE LEGACY OF EUCLID*, in *ZONING AND THE AMERICAN DREAM: PROMISES STILL TO KEEP* 101, 106–07 (Charles M. Haar & Jerold S. Kayden eds., 1999).

37. TROUNSTINE, *supra* note 18, at 112–18.

38. See generally James Gilbert Cassedy, *African Americans and the American Labor Movement*, 29 *FED. RECS. & AFR.-AM. HIST.* (1997); PHILIP S. FONER, *ORGANIZED LABOR AND THE BLACK WORKER* (1981).

39. KATZNELSON, *supra* note 20, at 113–41.

40. See Kevin E. Jason, *Dismantling the Pillars of White Supremacy: Obstacles in Eliminating Disparities and Achieving Racial Justice*, 23 *CUNY L. REV.* 139, 153–59 (2020) (describing policies initiated by Roosevelt administration's housing agencies and institutions that advance racial segregation, generated significant wealth in White communities, and concentrated poverty and other social ills in Black communities).

and created formal boundaries between communities that were experiencing or attempting integration.⁴¹

Though racial segregation was sanctioned and enforced through laws, it was equally shaped and perpetuated by the prejudicial actions and social customs of White Americans. Indeed, some scholars have argued that the network effects of racial segregation and discrimination engendered cartel-like conduct amongst “a range of all-White groups, like homeowners’ association, unions, school boards, local political parties, city councils, and other racially exclusive groups.”⁴² Those groups sought to preserve the race-based advantages that stemmed from exclusionary practices and policies.⁴³ For instance, legal scholar Daria Roithmayr observed that “[t]hese [all-White] groups gained significant social, economic and political profit—higher wages, higher property values, greater political power—from excluding on the basis of race.”⁴⁴ Initially, racially biased and socially exclusionary customs of private citizens were brazen and brutal and included lynchings, riots, and other incidents of targeted violence. But over time these practices and customs became more subtle and framed as matters of personal preference, freedom of choice, and individual rights. So, when White parents send their children to private schools,⁴⁵ White employers only hire from their social and professional networks, and White residents contest construction of public housing or homeless shelters in their neighborhoods, these decisions are no longer viewed as or are considered acts of self-segregation. In fact, these actions and

41. See, e.g., Johnny Miller, *Roads to Nowhere: How Infrastructure Built American Inequality*, THE GUARDIAN (Feb. 18, 2018, 2:30 AM), <https://www.theguardian.com/cities/2018/feb/21/roads-nowhere-infrastructure-american-inequality>.

42. Roithmayr, *supra* note 20, at 48.

43. *Id.*

44. *Id.*

45. See S. EDUC. FOUND., *A History of Private Schools & Race in the American South*, <https://www.southerneducation.org/publications/historyofprivateschools> (describing how private schools became a safe haven for Southern Whites seeking to keep their children in segregated schools following legal mandates to de-segregate public schools); Chris Ford, Stephenie Johnson & Lisette Partelow, *The Racist Origins of Private School Vouchers*, CTR. FOR AM. PROGRESS (July 12, 2017), <https://www.americanprogress.org/issues/education-k-12/reports/2017/07/12/435629/racist-origins-private-school-vouchers> (detailing how public funds are diverted to private schools through voucher programs, which in turn uphold racial segregation in schools); Matthew Di Carlo & Kinga Wysienska-Di Carlo, *Public and Private School Segregation in the District of Columbia*, ALBERT SHANKER INST. (2017), <https://www.shankerinstitute.org/resource/dcsegregation> (finding that contemporary racial segregation in Washington D.C. public schools is driven by White parents separating their children by sending them to private schools). See generally SEAN F. REARDON & JOHN T. YUN, PRIVATE SCHOOL RACIAL ENROLLMENTS AND SEGREGATION, in PUBLIC SCHOOL CHOICE VS. PRIVATE SCHOOL VOUCHERS (2003) (finding that White students are more racially isolated in private schools than public schools).

normative stances conveniently ignore the current state of American society, described as follows:

A typical white person lives in a neighborhood that is 75 percent white and 8 percent African American, while a typical African American person lives in a neighborhood that is only 35 percent white and 45 percent African American. . . In the United States, a low-income African American person is more than three times more likely to live in a neighborhood with a poverty rate of 40 percent or more than a white person is, and a low-income Latino person is more than twice as likely to live in such a neighborhood. These statistics show that racial residential segregation and racialized concentrated poverty persist today.⁴⁶

Together, these implicit actions, practices, and policies have become both normalized and distorted, and as a result, many people in the United States rationalize the current state of racial segregation and contemporary racial inequality as isolated aberrations rather than norms that require systemic change and actions by government and society.⁴⁷

B. A REVIEW OF THE SOCIAL, POLITICAL, AND EPISTEMIC IMPLICATIONS OF RACIAL SEGREGATION

Scholarship across several disciplines has advanced our understanding of the variegated consequences of racial segregation. Though its intention was physical separation, racial segregation also produced a number of social, economic, political, cultural, psychological, epistemic, intergenerational, and other consequences for both dominant and minority racial groups.⁴⁸ As

46. Solomon Greene, Margery Austin Turner & Ruth Gourevitch, *Racial Residential Segregation and Neighborhood Disparities*, URB. INST. (2017), <https://furtheringfairhousing.mit.edu/sites/default/files/documents/racial-residential-segregation-and-neighborhood-disparities.pdf> (citing PAUL JARGOWSKY, CENTURY FOUND., ARCHITECTURE OF SEGREGATION: CIVIL UNREST, THE CONCENTRATION OF POVERTY, AND PUBLIC POLICY (2015) (comparing the typical demographics of neighborhoods a White person and African American lives in); JOHN LOGAN & BRIAN STULTS, THE PERSISTENCE OF SEGREGATION IN THE METROPOLIS: NEW FINDINGS FROM THE 2010 CENSUS, PROJECT US 2010 (2011) (dictating where a low-income African American person lives compared to a low-income Latinx person)).

47. See Barbara Tomilson, *Powerblind Intersectionality: Feminist Revanchism and Inclusion as a One-Way Street*, in SEEING RACE AGAIN: COUNTERING COLORBLINDNESS ACROSS THE DISCIPLINES 175 (Kimberle Williams Crenshaw, Luke Charles Harris, Daniel Martinez HoSang & George Lipsitz eds., 2019) (explaining that colorblindness stems from an accepted norm that creates an illusory world where racism doesn't exist until an isolated event occurs in an individual's life).

48. E.g., Patrick Sharkey, *The Intergenerational Transmission of Context*, 113 AM. J. SOCIOLOGY 931 (2008) (noting how racial and economic inequalities in neighborhoods have endured throughout numerous generations).

discussed in Section II.A and illustrated in decades of multidisciplinary scholarship,⁴⁹ the primary drivers of racial segregation are not only governmental actions and policies but also individual and collective actions of White Americans and “all-White groups.” Therefore, it is important to examine some aspects of these collateral consequences of racial segregation on White Americans and then evaluate the impact of racial segregation on technology development and outcomes. This Section reviews some of the social, political, and epistemic implications because they are relevant for evaluation of algorithmic design, analysis, and outcomes.

Racial segregation causes social and spatial isolation of one racial group from other races, and over time this separation can become self-reinforcing because the distance and isolation caused by racial segregation breeds differentiation of and indifference to the “others.” Indeed, sociologist Eduardo Bonilla-Silva argues that subjecting several generations of White Americans to “high levels of social and spatial segregation and isolation from minorities” created what Bonilla-Silva labels as “‘white *habitus*,’ a racialized, uninterrupted socialization process that *conditions* and *creates* whites’ racial taste, perceptions, feelings, and emotions and their views on racial matters.”⁵⁰ Bonilla-Silva adds that a central consequence of “white habitus” is its promotion of “a sense of group belonging (a White culture of solidarity) and negative views about nonwhites.”⁵¹ During the eighteenth, nineteenth and twentieth centuries, these negative views of racial group differences were

49. See, e.g., ROTHSTEIN, *supra* note 12; TROUNSTINE, *supra* note 18. See generally ARNOLD HIRSCH, MAKING OF THE SECOND GHETTO (1993) (noting how the emerging White population in Chicago’s South Side area contributed to the city’s racial and housing segregation problems); KENNETH JACKSON, CRABGRASS FRONTIER: THE SUBURBANIZATION OF THE UNITED STATES (1985) (describing how American residential patterns are a result of a combination of social history with economic factors); SAMUEL KELTON ROBERTS JR., INFECTIOUS FEAR: POLITICS, DISEASE, AND THE HEALTH EFFECTS OF SEGREGATION (2009) (arguing how White politicians, during the tuberculosis crisis in the United States, promoted racially segregated policies in order to control the spread of the disease); JUDITH R. BLAU, RACE IN THE SCHOOLS: PERPETUATING WHITE DOMINANCE? (2003) (looking at how policies within the public school system in the United States reproduced advantages for only White students); MELVIN OLIVER & THOMAS M. SHAPIRO, BLACK WEALTH/WHITE WEALTH: A NEW PERSPECTIVE ON RACIAL INEQUALITY (1996) (analyzing the differences in how White and Black populations in the United States have accumulated wealth and how racial segregation has impacted such structures); CLAUD ANDERSON, BLACK LABOR, WHITE WEALTH: THE SEARCH FOR POWER AND ECONOMIC JUSTICE (1994) (examining the repercussions of Jim Crow policies on Black Americans); LAWRENCE T. BROWN, THE BLACK BUTTERFLY: THE HARMFUL POLITICS OF RACE AND SPACE IN AMERICA (2021) (probing how historic and current regulatory policies have shaped the hypersegregation issue in Baltimore, Maryland).

50. BONILLA-SILVA, *supra* note 20, at 152.

51. *Id.*

advanced through scientific racism theories and narratives.⁵² Even though these theories and narratives were ultimately challenged and discredited, notions of racial difference lived on because they were “built into the urban and suburban fabric” of American society and institutionalized in rules and laws.⁵³ Thus, socially constructed categories of difference that mattered in the past, in this case race, continue to matter and shape one’s world views, interests, and actions.⁵⁴

When thinking about race as socially constructed categories of differences, “whiteness is not perceived as a racial category, other categories are.”⁵⁵ This perception results in the false presumption that Whiteness is a norm, and therefore homogenous White neighborhoods are not products of racial segregation but rather “normal” neighborhoods.⁵⁶ Treating Whiteness as a norm in a racially diverse yet structurally unequal society is problematic because political, social, and economic activities and outputs are then designed with only White individuals and groups in mind. Those who are not White are

52. See DOROTHY E. ROBERTS, *FATAL INVENTION: HOW SCIENCE, POLITICS, AND BIG BUSINESS RE-CREATE RACE IN THE TWENTY-FIRST CENTURY* (2011) (examining scientific racism and how contemporary science and technologies may advance its logics and promote inequality); see also ANGELA SAINI, *SUPERIOR: THE RETURN OF RACE SCIENCE* (2019) (examining the history and re-emergence of scientific racism theories of the nineteenth century); Aaron Hanlon, *The Use of Dubious Science to Defend Racism is as Old as the Founding Fathers*, NBC NEWS (Nov. 25, 2017), <https://www.nbcnews.com/think/opinion/use-dubious-science-defend-racism-old-founding-fathers-ncna823116> (describing how scientific racism was central to eighteenth century Enlightenment thinking).

53. See Clarissa Rile Hayward, *Urban Space and American Political Development: Identity, Interest, Action*, in *THE CITY IN AMERICAN POLITICAL DEVELOPMENT* 141, 144 (Richardson Dilworth ed., 2009).

54. *Id.*; see also Kelly M. Hoffman, Sophie Trawalter, Jordan R. Axt & M. Norman Oliver, *Racial Bias In Pain Assessment and Treatment Recommendations, and False Beliefs About Biological Differences Between Blacks and Whites*, 113 *PROC. NAT’L ACAD. SCI.* 4296 (2016) (finding that seventy-three percent of White medical students wrongly believed Black people have a higher pain tolerance than White people); Jill Sheridan, *A Black Woman Says She Had to Hide Her Race to Get A Fair Home Appraisal*, NPR (May 21, 2021), <https://www.npr.org/2021/05/21/998536881/a-black-woman-says-she-had-to-hide-her-race-to-get-a-fair-home-appraisal> (describing how Black-owned homes are undervalued when compared to White-owned homes); see generally Jill D. Weinberg & Laura Beth Nielsen, *Examining Empathy: Discrimination, Experience, and Judicial Decisionmaking*, 85 *U. SOUTHERN CAL. L. REV.* 313 (2012) (finding that White federal judges dismiss employment discrimination cases involving non-White plaintiffs at higher rates than non-White judges, and non-White judges tend to assess discrimination claims differently than White judges).

55. BEVERLY DANIEL TATUM, *WHY ARE ALL THE BLACK KIDS SITTING TOGETHER IN THE CAFETERIA?: AND OTHER CONVERSATIONS ABOUT RACE* 93 (1997).

56. *Id.*

excluded and “seem not to fit because of something in their own nature.”⁵⁷ This perception has several repercussions, but this Article will focus on two of them for brevity.

First, treating Whiteness as the norm “begets a politics of parochial self-interest” because all-White groups are motivated to “maximize benefits for their own community and to limit fiscal burdens by denying access to populations and land uses that they perceive as undesirable.”⁵⁸ As a result, resources and benefits (e.g., higher property values, well-funded schools, newer amenities) concentrate in “White spaces,”⁵⁹ and collective or societal problems (e.g., poverty, over-policing, underfunded schools, higher taxes) accumulate in non-White spaces.⁶⁰ These racialized concentrations become worse because they “[occur] in an economic system that increasingly rewards the same affluent, professional, largely suburban class, creating gaps of opportunity that are unlikely ever to be closed.”⁶¹ This investment in a winner-take-all system also makes one less interested in systemic evaluations, reforms, or remedies. In fact, economist Glenn C. Loury argues that this is not just a political stance but an epistemic one.⁶² Loury calls it a “biased social cognition,” and defines it as “a politically consequential cognitive distortion to ascribe the disadvantage to be observed among a group of people to qualities thought to be intrinsic to that group when, in fact, that disadvantage is the product of a system of social interactions.”⁶³ As a result, White Americans develop “powerful explanations—which have ultimately become

57. MARTHA MINOW, MAKING ALL THE DIFFERENCE: INCLUSION, EXCLUSION, AND AMERICAN LAW 21 (1990); *see generally* Elijah Anderson, “*The White Space*,” 1 SOCIO. RACE & ETHNICITY 10 (2015) (noting the exclusionary nature of predominately White neighborhoods and communities and how Black people are forced to navigate such spaces as a condition of their existence).

58. Sheryll D. Cashin, *Drifting Apart: How Wealth and Race Segregation Are Reshaping the American Dream*, 47 VILL. L. REV. 595, 600 (2002); *see also* Hayward, *supra* note 53 (arguing that people who are privileged by extant power relations have and act on self-interested politics that concentrates societal problems in other, less privileged communities rather than their own).

59. Here, I am adopting the term “White space” as framed and employed by Sociologist Elijah Anderson. He refers to “the White space” as a perceptual category for “overwhelmingly White neighborhoods, restaurants, schools, universities, workplaces, churches and other associations, courthouses, and cemeteries, a situation that reinforces a normative sensibility in settings in which black people are typically absent, not expected, or marginalized when present.” Anderson, *supra* note 57, at 10. Anderson also notes that White people tend to regard “the White space” as “unremarkable, or as normal, taken-for-granted reflections of civil society.” *Id.*

60. HAYWARD, *supra* note 53, at 148–50; *see also* Cashin, *supra* note 58, at 595.

61. Cashin, *supra* note 58, at 596.

62. GLENN C. LOURY, THE ANATOMY OF RACIAL INEQUALITY 26 (2003).

63. *Id.*

justifications—for contemporary racial inequality that exculpate them from any responsibility for the status of people of color.”⁶⁴ This is why, despite the fact that Americans are growing increasingly aware of racial inequalities in the United States,⁶⁵ according to a 2020 Pew Research Center survey, the majority of White Americans believe the country has made enough progress on racial equality for Black people, and fifteen percent of White Americans believe this progress has gone too far.⁶⁶

Second, treating Whiteness as a norm stigmatizes racial difference because “problems of inequality can be exacerbated both by treating members of minority groups the same as members of the majority and by treating the two groups differently.”⁶⁷ Treating different racial groups in the same way means ignoring their structurally unequal positions, and possibly entrenching an unfair status quo (particularly social, political, and economic arrangements) rather than acknowledging inequality as a “part of the discriminating framework that must itself be changed.”⁶⁸ Applying color-blind logics, approaches, or tactics to color-bound problems also renders White racial dominance (i.e., the dominant social, political, and economic position of White Americans) invisible and can further entrench social inequalities across all racial groups.⁶⁹ And to be clear, acknowledging White racial dominance does not mean ignoring the deep class schisms amongst White Americans. Indeed, legal scholar Ian Haney-Lopez clarifies that:

64. BONILLA-SILVA, *supra* note 20, at 2; *see also* Charles W. Mills, *White Ignorance*, in RACE AND EPISTEMOLOGIES OF IGNORANCE 11, 28 (Shannan Sullivan & Nancy Tuana eds., 2007) (“[W]hite normativity underpins White privilege, in the first case by justifying differential treatment by race and in the second case by justifying formally equal treatment by race that—in its denial of the cumulative effect of past differential treatment—is tantamount to continuing it.”).

65. Katanga Johnson, *U.S. Public More Aware of Racial Inequality but Still Rejects Reparations: Reuters/Ipsos Polling*, REUTERS (June 25, 2020, 4:03 AM), <https://www.reuters.com/article/us-usa-economy-reparations-poll/u-s-public-more-aware-of-racial-inequality-but-still-rejects-reparations-reuters-ipsos-polling-idUSKBN23W1NG>.

66. Juliana Menasce Horowitz, Kim Parker, Anna Brown & Kiana Cox, *Amid National Reckoning, Americans Divided on Whether Increased Focus on Race Will Lead to Major Policy Change*, PEW RSCH. CTR. (Oct. 6, 2020), <https://www.pewresearch.org/social-trends/2020/10/06/amid-national-reckoning-americans-divided-on-whether-increased-focus-on-race-will-lead-to-major-policy-change>.

67. MINOW, *supra* note 57, at 20.

68. *Id.* at 76. *See generally* DERRICK BELL, AND WE ARE NOT SAVED: THE ELUSIVE QUEST FOR RACIAL JUSTICE (1987) (detailing the limitations of current thought processes for overcoming racial inequality issues); GEORGE LIPSITZ, HOW RACISM TAKES PLACE (2011) (arguing that racism exists because of current practices that alter opportunities along racial lines).

69. *See* Tomilson, *supra* note 47.

Rather than belying the power of race, however, these internal rifts more likely reflect race's utility in palliating intra-group conflict among whites. Racial ideology does not guarantee equality among whites; it serves rather to mask and distract from gross inequalities that divide that group. That said, it remains the case that whites as a race (though not all whites individually) have maintained their position at the social and material apogee for centuries...[and] being White affords advantages across the range of material and status divisions that mar our society.⁷⁰

Conversely, treating racial groups differently follows a tendency to “shoehorn the United States’ racial history into a rhetorically powerful but analytically crude [and shallow] story of ‘two societies,’” where “particular inequalities that appear statistically as ‘racial’ disparities are in fact embedded in multiple social relations and . . . the dominant modes of approaching this topic impede the understanding of this larger picture.”⁷¹ Under this approach, people are treated as monoliths rather than individuals, and such treatment not only erases other aspects of their identity (e.g., gender, ethnicity, ability, class, religion, sexual orientation), but it also reduces the complex relationships and dynamics across society to a one dimensional problem.⁷² Moreover, treating racial groups differently perpetuates racial difference and therefore reinforces unstated norms, in this case Whiteness, as well as historical or existing social arrangements (i.e., racial segregation).⁷³ This reinforcement in turn suggests that the status quo is neutral and natural, and again impedes systemic evaluations, reforms, or remedies.⁷⁴

The analysis provided in this Section is intentionally bound and will be expounded upon in future scholarship, but this outline provides sufficient context for what will be explored in the subsequent Sections.

70. IAN HANEY LOPEZ, *WHITE BY LAW: THE LEGAL CONSTRUCTION OF RACE*, 148–50 (10th Anniversary ed. 2006).

71. Adolph L. Reed & Merlin Chowkwanyun, *Race, Class, Crisis: The Discourse of Racial Disparity and its Analytical Discontents*, 48 *SOCIALIST REG.* 149, 150–51 (2012).

72. See PATRICIA HILL COLLINS, *BLACK FEMINIST THOUGHT: KNOWLEDGE, CONSCIOUSNESS, AND THE POLITICS OF EMPOWERMENT* (2000) (expressly exploring the intersectionality of being Black and female); Kimberle Crenshaw, *Demarginalizing the Intersections of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics*, 1 *U. CHI. LEGAL F.* 139, 140 (1989); GRACE LEE BOGGS & SCOTT KURASHIGE, *THE NEXT AMERICAN REVOLUTION: SUSTAINABLE ACTIVISM FOR THE TWENTY-FIRST CENTURY* 64 (2012).

73. MINOW, *supra* note 57, at 53.

74. *Id.*

III. HOW RACIAL SEGREGATION SHAPES DATA-DRIVEN TECHNOLOGIES AND ALGORITHMIC BIAS

Part III explores two aspects of how racial segregation impacts data-driven technologies and algorithmic bias. Section III.A examines how racial segregation and the societal inequality it breeds influence algorithmic design and analysis. Then, Section III.B considers how the failure to adequately assess the role of racial segregation and its consequences lead to the implementation of some data-driven technologies with minimal scrutiny and falsely positive evaluation.

A. RACIAL SEGREGATION AND ALGORITHMIC DESIGN

There is growing recognition that data-driven techniques and technologies, like predictive analytics and actuarial assessment, can produce racially discriminatory outcomes.⁷⁵ But there is less consensus on what drives these discriminatory outcomes. In particular, what component of the algorithmic design process is at fault and does context or the type of technology or technique matter? Within legal scholarship, there are two primary assertions. Some scholars point to human bias, error, or intervention as a source,⁷⁶ whereas others suggest that discrimination is an artifact of the data sources and data mining processes.⁷⁷ Yet, these assertions are not mutually exclusive. In

75. See, e.g., EXEC. OFFICE OF THE PRESIDENT, *BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES* (2014) (acknowledging a potential for discriminatory outcomes of predictive analytics); EUR. COMM'N, *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, SEC (2021) 167 final (Apr. 21, 2021).

76. E.g., Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 4 (2014) (“Because human beings program predictive algorithms, their biases and values are embedded into the software’s instructions.”); Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1254 (2008) (“Programmers routinely change the substance of rules when translating them from human language into computer code. The resulting distorted rules effectively constitute new policy that can affect large numbers of people.”).

77. E.g., Joshua A. Kroll, Joanna Huey, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson & Harlan Yu, *Accountable Algorithms*, 165 U. PA. L. REV. 633, 680 (2017) (“[A]lgorithms that include some type of machine learning can lead to discriminatory results if the algorithms are trained on historical examples that reflect past prejudice or implicit bias”); Solon Barocas & Andrew D. Selbst, *Big Data’s Disparate Impact*, 104 CALIF. L. REV. 671, 674 (2016) (“Discrimination may be an artifact of the datamining process itself, rather than a result of programmers assigning certain factors inappropriate weight.”); Rashida Richardson, Jason Schultz & Kate Crawford, *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, 94 N.Y.U. L. REV. 15, 41 (2019) (“[I]t becomes clear that any predictive policing system trained on or actively using data from jurisdictions with proven problematic conduct cannot be relied on to produce valid results”).

fact, accepting racial segregation as a root cause and source of algorithmic bias reveals that these assertions are connected.

Racial segregation is both enduring and pervasive, such that it shapes most aspects of society and is reflected in most institutional arrangements.⁷⁸ At the same time, White Americans, who dominate the technology sector and are therefore notably associated with algorithmic design and analysis, often perpetuate, normalize, and overlook racial segregation. Thus, racial segregation inevitably influences and shapes data sources, the data mining processes, and human biases and practices in the technology development process. To illustrate this point, Section III.A.1 explicates how training data can be systemically biased by racial segregation. Section III.A.2 then explores how data-driven technologies can replicate or amplify inequalities because the lack of understanding and evaluation of racial segregation and its consequences can influence human choices and decisions in the technology development process.

1. Training Data

Many data-driven technologies, particularly those that incorporate machine learning, rely on training data.⁷⁹ Training data is typically composed of samples of historical observations or curated examples considered relevant to performing a particular task (e.g., prediction or matching). Developers often classify training data into categories to train algorithms to behave in a certain way or produce specific outcomes.⁸⁰ As a result, training data can become systematically biased and contribute to algorithmic bias. Since it consists of historical samples or examples, training data can reflect social and structural inequalities in society. Human developers then calcify these inequalities through interventions or the lack thereof, particularly through decisions that generally classify or surmise the validity and appropriateness of particular data points or datasets.⁸¹

Police crime data, which is a primary data source for data-driven technologies used in policing, exemplifies how racial segregation can

78. See, e.g., George Lipsitz, *The Sounds of Silence*, in SEEING RACE AGAIN: COUNTERING COLORBLINDNESS ACROSS THE DISCIPLINES 23, 47 (Kimberle Williams Crenshaw, Luke Charles Harris, Daniel Martinez HoSang & George Lipsitz eds., 2019) (“The academy is a product of the society it studies . . . [i]t should not be a surprise that the pervasive patterns of segregation and subordination that shape society are evaded, ignored, or disavowed by colorblind constructs in history, law, education, economics, psychology, sociology, and urban planning.”).

79. Barocas & Selbst, *supra* note 77, at 680–81.

80. *Id.*; BEN GREEN, *THE SMART ENOUGH CITY* 66–69 (2019).

81. See Barocas & Selbst, *supra* note 77, at 680–81; GREEN, *supra* note 80, at 66–69; Sarah Brayne, *Big Data Surveillance: The Case of Policing*, 82 AM. SOCIO. REV. 977, 996–1004 (2017).

systematically bias training data. Extant research identifies racial residential segregation and the discriminatory public policies it enabled and, in many ways, concealed (e.g., uneven funding of public goods and divestment), as prominent structural forces in the reproduction of neighborhoods and geographies that are structurally unequal across racial and ethnic lines.⁸² And, as discussed in Part II, the confinement and concentration of societal problems and disadvantages—such as poverty, unemployment, and crime—in lower-income, non-White neighborhoods were consequences of racial segregation.⁸³ This geographic concentration of disadvantages has been linked to higher crime rates, though there is a debate regarding the causes of this correlation. Some scholars argue that the causal link is criminogenic factors inherent to certain disadvantaged communities,⁸⁴ whereas others believe that the cause is the lack of political power and organization to “implement strategies to improve social and institutional structures that affect crime.”⁸⁵ Yet, one fact remains true regardless of the causal explanation: sites of the concentrated disadvantages and social problems are often considered sites of “disorder,” where a greater degree of law enforcement presence, targeting, and surveillance practices are justified.⁸⁶ This law enforcement approach is colloquially known as “broken windows” or “hot spot” policing.⁸⁷

82. See Lauren J. Krivo, Ruth D. Peterson & Danielle C. Kuhl, *Segregation, Racial Structure, and Neighborhood Violent Crime*, 114 AM. J. SOCIO. 1765, 1768 (2009); see also JOHN R. LOGAN & HARVEY L. MOLOTCH, *URBAN FORTUNES: THE POLITICAL ECONOMY OF PLACE* (1987); DOUGLASS S. MASSEY & NANCY A. DENTON, *AMERICAN APARTHEID: SEGREGATION AND THE MAKING OF THE UNDERCLASS* (1993); Thomas L. McNutly, *The Residential Process and the Ecological Concentration of Race, Poverty, and Violent Crime in New York City*, 32 SOCIO. FOCUS 25 (1999); Gregory D. Squires & Charis E. Kubrin, *Privileged Places: Race, Uneven Development, and the Geography of Opportunity in Urban America*, 42 URB. STUD. 47 (2005); THOMAS SUGRUE, *THE ORIGINS OF THE URBAN CRISIS: RACE AND INEQUALITY IN POSTWAR DETROIT* (2016).

83. See MASSEY & DENTON, *supra* note 82; McNutly, *supra* note 82; Squires & Kubrin, *supra* note 82; SUGRUE, *supra* note 82.

84. See, e.g., MASSEY & DENTON, *supra* note 82; Douglas S. Massey, *Segregation and Violent Crime in Urban America*, in PROBLEMS OF THE CENTURY: RACIAL STRATIFICATION IN THE UNITED STATES 317 (Elijah Anderson & Douglas S. Massey eds., 2001).

85. Krivo et al., *supra* note 82, at 1771; see also Robert J. Sampson & William Julius Wilson, *Toward a Theory of Race, Crime, and Urban Inequality*, in CRIME AND INEQUALITY 37, 44–54 (John Hagan & Ruth D. Peterson eds., 1995).

86. See George L. Kelling & James Q. Wilson, *Broken Windows: The Police and Neighborhood Safety*, THE ATLANTIC, March 1982, <https://www.theatlantic.com/magazine/archive/1982/03/broken-windows/304465>; ANTHONY A. BRAGA & DAVID L. WEISBURD, *POLICING PROBLEM PLACES: CRIME HOT SPOTS AND EFFECTIVE PREVENTION* 35–60 (2010); RICHARD V. ERICSON & KEVIN D. HAGGERTY, *POLICING THE RISK SOCIETY* 39–80 (1997); see also BRIAN JEFFERSON, *DIGITIZE AND PUNISH: RACIAL CRIMINALIZATION IN THE DIGITAL AGE* 165–82 (2020) (describing the expansion of surveillance technologies to “deviant places”).

87. Kelling & Wilson, *supra* note 86; BRAGA & WEISBURD, *supra* note 86, at 45.

The result of this law enforcement approach is that the data police produce and use will reflect “hot spot” policing practices and policies.⁸⁸ This discriminatory law enforcement practice distorts crime data because even when certain types of crime occur equally across a large geographic area, police crime data may not accurately reflect that reality.⁸⁹ Instead, crime data will reflect where police officers concentrate their time, because crimes that occur in heavily patrolled public places become more visible and thus are more likely to be recorded,⁹⁰ and resulting crime datasets will include places and people who had more contacts with law enforcement and will not reflect the actual crime occurrence rates.⁹¹

This practice of policing “disorder” creates two issues in the context of so-called “big data” policing, an approach where police rely on crime data or data-driven technologies to inform policing strategies and practices. First, this application of the “disorder” label classifies neighborhoods and spaces as criminogenic and subjects their residents to negative, differential treatment and adverse categorization. For instance, “individuals living in low-income, minority areas have a higher probability of their ‘risk’ being quantified than those in more advantaged neighborhoods.”⁹² Moreover, since police contact is the entry point into the criminal justice system, arrests and incarcerations can concentrate in these disadvantaged neighborhoods⁹³ thus making them even worse.⁹⁴ Second, policing “disorder” produces feedback-loop effects⁹⁵ because the skewed crime data justifies greater police presence in lower-income, non-White neighborhoods and subjects their denizens to potentially lesser constitutional protection.⁹⁶ In fact, research suggests that the systematic

88. See generally Richardson et al., *supra* note 77.

89. See, e.g., Aaron Shapiro, *Reform Predictive Policing*, 541 NATURE 458, 460 (2017), <https://www.nature.com/articles/541458a.pdf> (finding that despite evidence of an even dispersal of drug use across all Oakland, a GIS algorithm named PredPol would send officers mostly to non-White neighborhoods); Kristian Lum & William Isaac, *To Predict and Serve?*, 5 SIGNIFICANCE (Oct. 7, 2016), <https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00960.x> (explaining how unequal trust in police leads to unequal crime reporting).

90. See Troy Duster, *Pattern, Purpose, and Race in the Drug War*, in CRACK IN AMERICA: DEMON DRUGS AND SOCIAL JUSTICE 265 (C. Reinerman & H.G. Levine eds., 1997).

91. Richardson et al., *supra* note 77, at 40–46 (2019); GREEN, *supra* note 80, at 70–89.

92. Brayne, *supra* note 81, at 997.

93. See generally TODD R. CLEAR, IMPRISONING COMMUNITIES: HOW MASS INCARCERATION MAKES DISADVANTAGED NEIGHBORHOODS WORSE (2009).

94. E.g., Brayne, *supra* note 81, at 977; Richardson et al., *supra*, note 77 at 15.

95. Richardson et al., *supra* note 77, at 40–48 (describing the feedback loop effects of biased crime data).

96. See Andrew Guthrie Ferguson, *Crime Mapping and the Fourth Amendment: Redrawing “High-Crime Areas”* 63 U.C. HASTINGS L.J. 101, 131–40 (2011) (detailing how high-crime area designations affect courts’ analysis of the Fourth Amendment reasonable suspicion standard

effect of this self-reinforcing policing practice is a complex and sophisticated method police use to enforce racial segregation because the increased police presence and activity compound concentrated disadvantages and reinforce segregative practices (e.g., further depreciating home values or accelerating “White flight”).⁹⁷

In sum, racial segregation constrains and informs policing practices, policies, and strategies that in turn shape police-generated crime data commonly used as training data for data-driven police technologies.

2. Problem Formulation

Most parts of the technology development process involve some form of human intervention or discretion, including decisions of what technologies to develop, what problems are appropriate to address with data-driven methods, and what methods to employ. This Section uses a specific example to demonstrate how the failure to adequately understand racial segregation and its consequences in formulating data science problems (“problem formulation”) can negatively skew algorithmic outcomes. But first, it is important to understand the role of problem formulation in the technology development process.

Problem formulation is a foundational yet vexed part of the technology development process. It is foundational because it determines whether applying appropriate algorithms can resolve a societal or business problem.⁹⁸ It is vexed because it can influence most of developers’ design choices and decisions.⁹⁹ Problem formulation requires developers to engage in various forms of discretionary work to measure and translate complex and amorphous problems into formal terms that can be parsed and solved by algorithms.¹⁰⁰ Such discretionary work can include determining the nature of the problem, ways to measure the problem, concrete issues a developer seeks to solve or

and thus lower constitutional protection of those areas); *Leaders of a Beautiful Struggle v. Baltimore Police Dep’t*, 979 F.3d 219, 226–32 (4th Cir. 2020) (finding the use of an aerial surveillance program constitutional because it tracked only “short-term movements in public” and served a “critical government purpose” of combatting violent crime in a high-crime area).

97. Grace Roberts, Dylan Horwitz, Evan Horowitz & Cameron Tomaiko, *The American System: How Police Enforce Segregation*, ARCGIS STORYMAPS (Dec. 19, 2019), <https://storymaps.arcgis.com/stories/ac3d72c7b1c54305937e40d2ad43d774>.

98. Nicole Scott, *Defining A Data Science Problem*, TOWARDS DATA SCI. (Aug. 18, 2019), <https://towardsdatascience.com/defining-a-data-science-problem-4cbf15a2a461>.

99. See generally Samir Passi & Solon Barocas, *Problem Formulation and Fairness*, PROCS. ACM CONF. ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY (2019) (explaining how problem formulation influences goals and outcomes).

100. *Id.*; Barocas & Selbst, *supra* note 77, at 678.

understand, metrics for success, and methods that are best suited to solve the problem or perform a certain task (e.g., prediction or allocation).

School assignment algorithms are an instructive example of how developers' presumptions, oversights, and choices in problem formulation can contribute to algorithmic bias and antithetical outcomes. School assignment algorithms are typically centralized algorithms used to assign K–12 students to specific schools based on families' ranked preferences.¹⁰¹ In 2012, Boston Public Schools adopted a school assignment algorithm to “reengineer its school choice and assignment system, with the goal of providing parents with equitable access to good schools that are close to home,”¹⁰² but the algorithm “largely maintained status quo, inheriting but not counteracting inequities that existed under the previous system . . .”¹⁰³ The goals for implementing the algorithm were ambitious: to increase students' access to high-quality schools while reducing the distance students travel to get to school.¹⁰⁴ Yet, some outcomes were unsatisfactory: for example, reducing racial and geographic integration across the school district.¹⁰⁵ These outcomes stemmed from the problem formulation phase, during which the developers of the school assignment algorithm demonstrated a fairly myopic understanding of racial segregation in public schools and how it relates to school quality. This myopic understanding constrained the problem formulation analysis and overall expectations.

The developers of the algorithm did acknowledge that their chosen mathematical approach to assigning students from particular communities to less geographically dispersed schools, a so-called “correlated lottery,” “may cause racial or socio-economic segregation, because race and socio-economic status are correlated with geography.”¹⁰⁶ Yet, the developers failed to account for reasons behind this correlation and their effect on school quality and

101. Matt Kasman & Jon Valant, *The Opportunities and Risks of K-12 Student Placement Algorithms*, BROOKINGS INST. (Feb. 28, 2019), <https://www.brookings.edu/research/the-opportunities-and-risks-of-k-12-student-placement-algorithms>.

102. BOS. AREA RSCH. INITIATIVE, AN EVALUATION OF EQUITY IN THE BOSTON PUBLIC SCHOOLS' HOME-BASED ASSIGNMENT POLICY 10 (2018), <https://www.bostonpublicschools.org/cms/lib/MA01906464/Centricity/Domain/162/Final%20Evaluation%20of%20Equity%20in%20BPS%20HBAP.pdf>.

103. *Id.* at 70.

104. *Id.* at 1.

105. *Id.* at 3.

106. ITAI ASHLAGI & PENG SHI, IMPROVING COMMUNITY COHESION IN SCHOOL CHOICE VIA CORRELATED-LOTTERY IMPLEMENTATION ec6 (2014), <http://web.mit.edu/iashlagi/www/papers/correlated-lottery.pdf>.

locations of “good” schools.¹⁰⁷ Racial and socioeconomic segregation in public schools is both a byproduct of residential segregation (including concentrated disadvantage) and discriminatory public policies (e.g., school funding and school district mapping).¹⁰⁸ These problems directly impact school quality and the locations of high-quality schools,¹⁰⁹ and the failure to understand these root causes and dynamics constrained the developers’ problem formulation. Indeed, an evaluation of the school assignment algorithm’s outcomes found that disparities in access to high quality schools reproduced by the algorithm directly reflected the uneven geographic distribution of school quality in Boston.¹¹⁰ Thus, in this example, the developers of the school assignment algorithm failed to understand and adequately evaluate how racial segregation affects school quality and the geographic distribution of quality schools, and this failure hindered the problem formulation analysis that would otherwise reveal the futility of the chosen approach.¹¹¹

107. The definition of a “good school” is normatively based and can often depend on intangible or hard to define qualities. In public and academic discourse, this term often refers to sufficient funding, optimal class size, experienced teachers, and availability of academic and extracurricular programs. *See, e.g.*, Dwyer Gunn, *Non-White School Districts Get \$23 Billion Less Funding Than White Ones*, PAC. STANDARD MAG. (Feb. 26, 2019), <https://psmag.com/education/nonwhite-school-districts-get-23-billion-less-funding-than-white-ones>; ERICA FRANKENBERG, JONGYEON EE, JENNIFER B. AYSUCUE & GARY ORFIELD, *HARMING OUR COMMON FUTURE: AMERICA’S SEGREGATED SCHOOLS 65 YEARS AFTER BROWN 23–24* (2019), <https://civilrightsproject.ucla.edu/research/k-12-education/integration-and-diversity/harming-our-common-future-americas-segregated-schools-65-years-after-brown/Brown-65-050919v4-final.pdf>; GARY ORFIELD & CHUNGMEI LEE, *WHY SEGREGATION MATTERS: POVERTY AND EDUCATIONAL INEQUALITY 7* (2005), <https://civilrightsproject.ucla.edu/research/k-12-education/integration-and-diversity/why-segregation-matters-poverty-and-educational-inequality/orfield-why-segregation-matters-2005.pdf>.

108. *See, e.g.*, Gunn, *supra* note 107; Alvin Chang, *We Can Draw School Zones to Make Classrooms Less Segregated. This Is How Well Your District Does.*, VOX (Aug. 27, 2018), <https://www.vox.com/2018/1/8/16822374/school-segregation-gerrymander-map>; FRANKENBERG ET AL., *supra* note 107, at 23–24.

109. *See* Richard Rothstein, *The Racial Achievement Gap, Segregated Schools, and Segregated Neighborhoods: A Constitutional Insult*, 7 RACE & SOC. PROBS. 21, 21–22 (2015); *see also* GARY ORFIELD & CHUNGMEI LEE, C.R. PROJECT, *WHY SEGREGATION MATTERS: POVERTY AND EDUCATIONAL INEQUALITY* (2005) (noting how socioeconomic segregation is a cause of educational inequality).

110. BOS. AREA RSCH. INITIATIVE, *supra* note 102, at 2 (“The overarching lesson of the evaluation is that it is impossible for a choice and assignment system to provide access to ‘good schools close to home’ when the geographic distribution of quality schools is itself inequitable.”).

111. *See* Roel Dobbe, Thomas Krendl Gilbert, & Yonatan Mintz, *Hard Choices in Artificial Intelligence*, 300 A. I. 1, 9 (2021) (stating that AI system developers must “remain attentive to the fundamental limitations of technical solutions to resolve them,” which means an

B. RACIAL SEGREGATION AND ALGORITHMIC EVALUATION

This Part explores how government officials adopt and researchers positively evaluate data-driven applications and technologies that are inextricably shaped by racial segregation, and how this entanglement is ignored or undervalued. To illustrate the problem, this Part focuses on crime-focused geographic information systems (GIS), a category of data-driven applications and technologies notable for their broad adoption and acceptance but deserving greater scrutiny.¹¹²

GIS technologies are computer-based tools that capture, store, analyze, query, and visualize geospatial data and related information.¹¹³ These tools often combine crime mapping and statistical analysis to inform law enforcement strategies and practices.¹¹⁴ Commonly used examples of GIS technologies include CompStat and predictive policing.¹¹⁵ Most GIS

“alertness” to “all factors responsible for a situation, including social and political components”).

112. See, e.g., Press Release, U.S. Senator Ron Wyden, Wyden, Democrats Question DOJ Funding of Unproven Predictive Policing Technology (Apr. 15, 2021), <https://www.wyden.senate.gov/news/press-releases/wyden-democrats-question-doj-funding-of-unproven-predictive-policing-technology> (stating eight Senators asked DOJ to halt funding of predictive policing programs until DOJ can ensure proper documentation, independent audits, and a system of due process for the impacted individuals); Joseph L. Giacalone & Alex S. Vitale, *When Policing Stats Do More Harm than Good: Column*, USA TODAY (Feb. 9, 2017), <https://www.usatoday.com/story/opinion/policing/spotlight/2017/02/09/compstat-computer-police-policing-the-usa-community/97568874> (highlighting how the NYPD’s overemphasis on crime statistics and GIS led to a numbers game instead of actual crime control).

113. U.S. DEP’T. OF HOMELAND SEC., GEOGRAPHIC INFORMATION SYSTEMS AND PREDICTIVE POLICING APPLICATION NOTE 1 (2013), https://www.dhs.gov/sites/default/files/publications/GIS-Predictive-Policing-AppN_0813-508_0.pdf; Ferguson, *supra* note 96, at 187.

114. Although GIS technologies are more prevalent within the law enforcement sector, they gain popularity in other sectors that use risk-based practices or strategies, such as child welfare. For example, a child welfare program “Predict-Align-Prevent” uses place-based geospatial machine learning to identify and target children and communities “at greatest risk of maltreatment.” PREDICT ALIGN PREVENT, RICHMOND, VIRGINIA TECHNICAL REPORT 4 (2019), https://b9157c41-5fbc-4e28-8784-ea36ffdbce2f.filesusr.com/ugd/fbb580_2f1dda2ff6b84f32856bc95d802d6629.pdf [hereinafter PREDICT ALIGN PREVENT, RICHMOND]. But the machine learning was based on prior maltreatment incident data and mapped maltreatment risk areas based on local child welfare, health, crime, code violations, and infrastructure data. See, e.g., *id.* at 8–10; PREDICT-ALIGN-PREVENT, LITTLE ROCK, ARKANSAS TECHNICAL REPORT FOR THE ARKANSAS DIVISION OF CHILDREN AND FAMILY SERVICES, <https://papreports.org/little-rock-ar/index.html> (last visited July 1, 2021).

115. See PREDICT ALIGN PREVENT, RICHMOND, *supra* note 114, at 4; see also U.S. DEP’T. OF HOMELAND SEC., *supra* note 113, at 9 (highlighting various predictive policing vendors); DAVID WEISBURD, STEPHEN D. MASTROFSKI, ROSANN GREENSPAN & JAMES J. WILLIS, NAT’L POLICE FOUND., THE GROWTH OF COMPSTAT IN AMERICAN POLICING 6 (2004),

technologies are predicated on criminology or policing theories that seek to optimize the policing of “disorder” by revealing spatial knowledge and patterns embedded in the data.¹¹⁶ But GIS technologies and their underlying theories are all premised on the common fallacy that data analyses, or in this case data visualizations (i.e., crime maps), are merely revealing spatial knowledge. In reality, GIS technologies and their outputs, like crime maps or “hot spot” analyses, create a specific version of spatial knowledge that is shaped by normative views of social space (and its denizens), disorder, and crime, as well as local political priorities.¹¹⁷ Indeed, critical geographic information science scholars emphasize that mapmaking and other GIS approaches or technologies “invariably offers distorted representations of social reality . . . [because] data production can only yield a small ‘selection from the sum total of all possible data available [and] as such, data are inherently partial [and] selective.’”¹¹⁸

For instance, GIS technologies target “street crime” (e.g., larceny, vandalism, burglary) most prominently and rely on related statistics,¹¹⁹ even though the individual victimization rates and overall societal costs are greater for offenses like cybercrime, communications fraud, or corporate crimes.¹²⁰

<https://www.policefoundation.org/publication/the-growth-of-compstat-in-american-policing> (finding that almost a half of police departments with 100 or more sworn officers have used or plans to use CompStat-like programs, and almost thirty percent of the departments with 50–99 sworn officers plan to implement such a program).

116. Ferguson, *supra* note 96, at 109–12; George L. Kelling & William J. Bratton, *Declining Crime Rates: Insiders’ Views of the New York City Story*, 88 J. CRIM. L. & CRIMINOLOGY 1217, 1218–20 (1998); Andrew Guthrie Ferguson, *Predictive Policing Theory*, in THE CAMBRIDGE HANDBOOK OF POLICING IN THE UNITED STATES 491, 503–04 (Tamara Rice Lave & Eric J. Miller eds., 2019); ELI B. SILVERMAN, *NYPD BATTLES CRIME: INNOVATIVE STRATEGIES IN POLICING* (1999).

117. See Brian Jordan Jefferson, *Policing, Data, and Power-geometry: Intersections of Crime Analytics and Race During Urban Restructuring*, 39 URB. GEOGRAPHY 1247, 1249–54 (2018); Richardson et al., *supra* note 77, at 22; Katherine McKittrick, *On Plantations, Prisons, and a Black Sense of Place*, 12 SOC. & CULTURAL GEOGRAPHY 947, 954 (2011).

118. Jefferson, *supra* note 117, at 1249–50; see also MATTHEW W. WILSON, *NEW LINES: CRITICAL GIS AND THE TROUBLE OF THE MAP* (2017); ROB KITCHIN, *THE DATA REVOLUTION: BIG DATA, OPEN DATA, DATA INFRASTRUCTURES AND THEIR CONSEQUENCES* 3 (2014). See generally Nadine Schuurman, *Formalization Matters: Critical GIS and Ontology Research*, 96(4) ANNALS ASSOC. AM. GEOGRAPHERS 726, 726–28 (2006) (describing evolution of critical GIS theories).

119. E.g., Ferguson, *supra* note 96, at 104 n.11; see, e.g., *Residential Burglaries: An Ounce of Prevention Is Worth a Pound of Prosecution*, GEOLITICA (Apr. 5, 2021), <https://geolitica.com/blog/residential-burglaries-ounce-prevention-worth-pound-prosecution> (describing how the Geolitica platform can help prevent residential burglaries).

120. See, e.g., RJ Reinhart, *One in Four Americans Have Experienced Cybercrime*, GALLUP (Dec. 11, 2018) (highlighting the prevalence and frequency of cybercrime); Press Release, Fed. Trade Comm’n, *New FTC Data Shows that FTC Received Nearly 1.7 Million Fraud Reports*, and

This distinction between crimes and their societal costs is significant not only because the latter category of crimes tend to lack geographic boundaries required for GIS approaches, but they also provoke a different cultural response that is inherently shaped by racial segregation. Although cybercrime, communications fraud, or corporate crimes may be associated with “White spaces” like financial districts or areas with greater internet or technology access,¹²¹ the lack of clear geographic boundaries means that these spaces are not racialized or equated with criminality and risk.¹²² Moreover, research demonstrates that the social isolation and financial advantage experienced by most White Americans as a result of racial segregation leads to crime-specific cultural frames. These frames produce a neutralized response or justifications for white-collar crime and crimes committed by Whites, whereas the stereotype of “black criminality” is seen as a race problem that evokes punitive responses or a lack of empathy.¹²³ As a result, GIS technologies are understood and adopted as neutral technologies that analyze crime data to proactively “implement more efficient targeted policing practices at the precinct level . . .

FTC Lawsuits Returned \$232 Million to Consumers in 2019 (2020), <https://www.ftc.gov/news-events/press-releases/2020/01/new-ftc-data-shows-ftc-received-nearly-17-million-fraud-reports> (detailing massive communications fraud in the United States); Giulio Saggin, *What If Street Crime Statistics Matched Those of Cybercrime?*, HACKERNOON (Nov. 12, 2019), <https://hackernoon.com/if-street-crime-statistics-matched-those-of-cybercrime-mayhem-would-ensue-7x1d3233> (highlighting the high occurrence rate and gravity of cybercrimes); Roomy Khan, *White-Collar Crimes—Motivations and Triggers*, FORBES (Feb. 22, 2018, 01:06 PM), <https://www.forbes.com/sites/roomykhann/2018/02/22/white-collar-crimes-motivations-and-triggers/?sh=665ce9401219> (explaining the prevalence of white-collar crimes and their societal costs).

121. See Brian Clifton, Sam Lavigne & Francis Tseng, *White Collar Crime Risk Zones*, NEW INQUIRY MAG. (Mar. 2017), <https://whitecollar.thenewinquiry.com>; S. Derek Turner, FREE PRESS, *Digital Denied: The Impact of Systemic Racial Discrimination on Home-Internet Adoption* 70–75 (2016), https://www.freepress.net/sites/default/files/legacy-policy/digital_denied_free_press_report_december_2016.pdf (demonstrating a stark racial divide in home-internet access, adoption, and use).

122. See, e.g., Saggin, *supra* note 120 (“Over 80% of US adults believe cybercrime should be treated as a criminal act, yet nearly 25% believe stealing information online is not as bad as stealing property in real life.”). See generally JAMES D. UNNEVER & SHAUN L. GABBIDON, *A THEORY OF AFRICAN AMERICAN OFFENDING: RACE, RACISM AND CRIME* (2011) (detailing a range of race-linked social forces and patterns of inequality that influence racial imbalances in crime and crime data); KHALIL GIBRAN MUHAMMAD, *THE CONDEMNATION OF BLACKNESS: RACE, CRIME, AND THE MAKING OF MODERN URBAN AMERICA* (2010) (detailing the history and practice of using crime statistics to create the racist myth of black criminality and how these practices have influenced racially biased urban development and social policies).

123. Tracy Sohoni & Melissa Rorie, *The Whiteness of White-Collar Crime in the United States: Examining the Role of Race in a Culture of Elite White-Collar Offending*, 25 THEORETICAL CRIMINOLOGY 66, 73 (2021). See generally MUHAMMAD, *supra* note 122 (describing how race and crime statistics were artificially linked to create a stereotype of black criminality).

[and] to monitor the effectiveness of different police practices.”¹²⁴ But in practice these technologies selectively focus on a small subset of crimes and locations and serve to mask or legitimize disproportionate and often discriminatory policing of certain areas or communities.¹²⁵

Despite this flawed premise, law enforcement officials, technology vendors, and even some scholars argue that GIS technologies have a positive impact on reducing crime,¹²⁶ are less problematic than person-focused data-driven technologies,¹²⁷ and are legally and socially permissible with reasonable safeguards or a cautious design.¹²⁸ Yet, in order to hold this position one must

124. Adam Benforado, *The Geography of Criminal Law*, 31 CARDOZO L. REV. 823, 860 (2010).

125. See Jefferson, *supra* note 117, at 1255–56; Aaron Shapiro, *Reform Predictive Policing*, 541 NATURE 458, 460 (2017), <https://www.nature.com/articles/541458a.pdf> (explaining how focusing on racially biased data sometimes justifies policing mostly non-White communities even for crimes that are statistically evenly spread amongst all communities); *United States v. Curry*, 965 F.3d 313, 344 (4th Cir. 2020) (en banc) (Thacker, J. concurring) (“Predictive policing is merely a covert effort to attempt to justify racial profiling. Over time, predictive policing has been shown to be, at best, of questionable effectiveness, and at worst, deeply flawed and infused with racial bias.”); Brief of Chicago Community-Based Organizations, Brighton Park Neighborhood Council, et al. as Amici Curiae Supporting Defendant, *Illinois v. Williams*, (No. 20 CR 0889601) (Cook Cnty. Cir. Ct. filed May 3, 2021), <https://endpolicesurveillance.com/documents/2021-05-03-Motion-for-Leave-to-File-Brief-as-Amici-Curiae-with-Ex.-A-Amicus-Brief-attached.pdf> (arguing that the City of Chicago only deployed ShotSpotter in police districts with the largest proportion of Black and Latinx residents and the technology justifies discriminatory over-policing).

126. E.g., WILLIAM BRATTON, *TURNAROUND: HOW AMERICA’S TOP REVERSE THE CRIME EPIDEMIC* (1998); see, e.g., Avi Asher-Schapiro, *In a U.S. First, California City Set to Ban Predictive Policing*, REUTERS (June 17, 2020), <https://www.reuters.com/article/us-usa-police-tech-trfn-idUSKBN23O31A> (“On its website, PredPol said that its technology helps police fight crime and that Santa Cruz police reported a 19% reduction in burglaries since implementing its programme while Los Angeles Police saw a 25% fall.”); G. O. Mohler, M. B. Short, Sean Malinowski, Mark Johnson, G. E. Tita, Andrea L. Bertozzi & P. J. Brantingham, *Randomized Controlled Field Trials of Predictive Policing*, 110 J. AM. STAT. ASS’N 1399, 1409–10 (2015) (praising the effectiveness of crime forecasting and predictive policing).

127. E.g., Andrew G. Ferguson, *Policing Predictive Policing*, 94 WASH. U. L. REV. 1109, 1143 (2017) (“Suspicion based on correlation may be acceptable when talking about place-based crimes, but it is insufficient when talking about person-based crime. Sending a police car to patrol a suspected area is less consequential than sending a police detective to interrogate a suspect.”); SHOTSPOTTER, *A Citizen’s Guide to ShotSpotter Connect 6* (2021), https://www.shotspotter.com/wp-content/uploads/2021/03/ConnectCitizensGuide_v1_0.pdf (“The team behind Connect has carefully thought about the data used by the model to reduce potential harm to community. As part of this approach, we do not make predictions about the actions of people . . .”).

128. E.g., Ferguson, *supra* note 96, at 141–45 (2011) (suggesting a modification to the Fourth Amendment reasonable suspicion standard analysis that would make certain uses of GIS constitutionally permissible); SHOTSPOTTER, *supra* note 127, at 7 (listing ShotSpotter’s

completely ignore racial segregation and its consequences, other concurrent social or policy changes,¹²⁹ and some of the aforementioned practical implications of GIS technologies.¹³⁰ To help illustrate this point and emphasize why racial segregation must be considered when evaluating GIS and other data-driven technologies, this Section highlights a specific use case in Chicago, Illinois. The case was considered a success, yet its supporters overlooked the role of racial segregation as the source of structural conditions that drove crime as well as an alternative solution to address crime.

In 2017, the Chicago Police Department (CPD) launched a nine million dollar project piloting several GIS technologies—ShotSpotter (a location-based gun detection system), HunchLab (a place-based predictive policing system now owned by ShotSpotter and called ShotSpotter Connect), and Police Observation Devices (a system of CCTV cameras).¹³¹ These technologies were integrated into police district-based intelligence hubs known as Strategic Decision Support Centers (SDSCs), where information was analyzed to inform police district practices and strategies to compliment traditional policing.¹³² The pilot was initially implemented in six police districts

features designed to reduce over-policing by keeping track of length and frequency of police visits to a particular area).

129. See, e.g., Steven D. Levitt, *Understanding Why Crime Fell in the 1990s: Four Factors that Explain the Decline and Six that Do Not*, 18 J. ECON. PERSPECTIVES 163, 172–73, 176–83 (2004) (challenging narratives that “hot spots” and “community policing” practices played an important role in sharply reducing crime in the 1990s, and highlighting other factors that contributed to the crime reduction: increases in the number of police, the rising prison population, the receding crack epidemic, and the legalization of abortion); John Eterno & Eli B. Silverman, *The NYPD’s Compstat: Compare Statistics or Statistics?*, 12 INT’L J. POLICE SCI. & MGMT. 426, 428–29, 442 (2010) (finding that the implementation of CompStat created institutional pressure on police commanders to keep crime index low, which led to “unethical distortion of crime reports” by improperly downgrading felonies to misdemeanors, undervaluing the property loss to crime, reporting series of crimes as a single event, etc.).

130. E.g., Beryl Lipton, *“It’s Predpol, and It’s Going to Reduce Crime”: Agencies Take Algorithmic Effectiveness on Faith, with Few Checks in Place*, MUCKROCK (Nov. 5, 2019), <https://www.muckrock.com/news/archives/2019/nov/05/predictive-policing-lacks-accuracy-tests> (stating that because PredPol data was generated by “questionable policing strategies” the system simply automated racial bias); Mitchell L. Doucette, Christa Green, Jennifer Necci Dineen, David Shapiro & Kerri M. Raissian, *Impact of ShotSpotter Technology on Firearm Homicides and Arrests Among Large Metropolitan Counties: A Longitudinal Analysis 1999–2016*, J. URB. HEALTH (2021) (finding that implementing ShotSpotter technology has no significant impact on gun-related homicides or arrest outcomes).

131. See Michael Wasny, *The Shots Heard Round the City*, S. SIDE WEEKLY (Dec. 19, 2017), <https://southsideweekly.com/shots-heard-round-city-shotspotter-chicago-police/>; Timothy McLaughlin, *As Shootings Soar, Chicago Police Use Technology to Predict Crime*, REUTERS (Aug. 5, 2017, 3:25 AM), <https://www.reuters.com/article/us-chicago-police-technology-idUSKBN1AL08P>.

132. See Wasny, *supra* note 131; McLaughlin, *supra* note 131.

and then expanded to six additional districts, all of which were targeted for their high gun violence and homicide rates.¹³³ Local government officials received the project positively, relayed their satisfaction to some of the technology vendors,¹³⁴ and attributed declining crime rates to the SDSCs.¹³⁵ Specifically, the officials noted a thirty-nine percent decline in shootings and a thirty-three percent drop in murders, while the citywide number of murders was up three percent.¹³⁶

Yet, local residents and some critics questioned whether the crime reduction can simply be attributed to the technologies or even the CPD and feared that “the technology could prove a distraction from confronting what they say are the underlying issues driving violence in the city of 2.7 million.”¹³⁷ And an understanding of the role of racial segregation warrants this doubt. First, racial segregation cannot and should not be ignored in a city like Chicago since it is considered one of the most racially and economically segregated cities in the United States.¹³⁸ In fact, a 2016 report declared that “some 72% of black or White residents would have to move to a different census tract to even out the numbers, according to a commonly used segregation measure called the index of dissimilarity. In New York, the figure is 65% and in Philadelphia, it’s 63%.”¹³⁹

Most of the neighborhoods selected for the pilot and especially those celebrated for crime reduction—Englewood, West Garfield Park, and Austin—are almost exclusively comprised of Black and Latino residents.¹⁴⁰

133. Wasny, *supra* note 131.

134. See Robert Cheetham, *Why We Sold HunchLab*, AZAVEA BLOG (Jan. 23, 2019), <https://www.azavea.com/blog/2019/01/23/why-we-sold-hunchlab> (explaining that developers of HunchLab sold it because “the product was gaining some traction” its biggest customer—the CPD—was “seeing some success with the software”).

135. Mclaughlin, *supra* note 131.

136. *Id.* (stating that “the number of shootings in the 7th District from January through July fell 39 percent compared with the same period last year. The number of murders dropped by 33 percent to 34. Citywide, the number of murders is up 3 percent at 402.”).

137. *Id.*; see also Wasny, *supra* note 131 (listing factors that potentially distorted the results attributed to ShotSpotter).

138. CENSUSSCOPE, US METRO AREAS RANKED BY WHITE/BLACK DISSIMILARITY INDEX, https://www.censusscope.org/us/rank_dissimilarity_white_black.html (last visited July 2, 2021).

139. Tami Luhby, *Chicago: American’s Most Segregated City*, CNN BUS. (Jan. 5, 2016, 4:26 AM), <https://money.cnn.com/2016/01/05/news/economy/chicago-segregated>.

140. The population of the Englewood neighborhood, covered by the 7th CPD police district, is 94.1% Black and 4.3% Latinx. CHICAGO METRO. AGENCY FOR PLANNING, COMMUNITY DATA SNAPSHOT ENGLEWOOD, CHICAGO COMMUNITY AREA JUNE 2021 RELEASE 3 (2020), <https://www.cmap.illinois.gov/documents/10180/126764/Englewood.pdf>. The population of West Garfield Park neighborhood, covered by the 11th CPD police district, is 93.7% Black and 2.6% Latinx. CHICAGO METRO. AGENCY FOR

These neighborhoods have substantially higher poverty rates than Chicago overall¹⁴¹ and were negatively impacted by segregation practices and policies such as the construction of the Dan Ryan Expressway, urban renewal projects (specifically the demolition or redevelopment of public housing), public and private divestment, and school closures.¹⁴² Research on hypersegregation demonstrates that racial minorities living in hypersegregated areas are exposed to concentrated disadvantages and distress, including elevated crime and violence.¹⁴³ Research on the societal costs of segregation finds that black-White segregation (not economic segregation) strongly correlates with higher homicide rates.¹⁴⁴ The CPD chose these locations to run the pilot because of

PLANNING, COMMUNITY DATA SNAPSHOT WEST GARFIELD PARK, CHICAGO COMMUNITY AREA JUNE 2021 RELEASE 3 (2020), <https://www.cmap.illinois.gov/documents/10180/126764/West+Garfield+Park.pdf> [hereinafter COMMUNITY DATA SNAPSHOT WEST GARFIELD PARK]. The population of the Austin neighborhood, covered by the 15th CPD police district, is 79.1% Black and 14.4% Latinx. CHICAGO METRO. AGENCY FOR PLANNING, COMMUNITY DATA SNAPSHOT AUSTIN, CHICAGO COMMUNITY AREA JUNE 2021 RELEASE 3 (2020), <https://www.cmap.illinois.gov/documents/10180/126764/Austin.pdf>.

141. The Englewood neighborhood, covered by the 7th CPD police district, has a poverty rate of forty-four percent, while the overall Chicago poverty rate is twenty percent. METRO. PLANNING COUNCIL, DEMOGRAPHICS—CHICAGO (2009), <https://www.metroplanning.org/uploads/cms/documents/olympicsenglewooddemographics.pdf>. The West Garfield Park neighborhood, covered by the 11th CPD police district, has a median household income of \$24,591, while Chicago's median household income is \$55,198. COMMUNITY DATA SNAPSHOT WEST GARFIELD PARK, *supra* note 140, at 5.

142. See Dahleen Glanton, *Lingering Lines of Discrimination*, CHICAGO TRIBUNE (Mar. 1, 1998), <https://www.chicagotribune.com/news/ct-xpm-1998-03-01-9803010173-story.html> (describing how businesses refuse to provide services in “high-risk” neighborhoods or charge premium prices if they have to); Scott Smith, *The Intersection of the Dan Ryan and Chicago Segregation*, OUR MAN IN CHICAGO (Apr. 18, 2021), <http://www.ourmaninchicago.net/2021/04/the-intersection-of-the-dan-ryan-and-chicago-segregation> (explaining how building of the Dan Ryan Expressway prevented Black people's expansion to White neighborhood and “helped expedite the exodus of the white community”); Alvin Ulido Lumbanraja, *To Kill a Neighborhood: Urban Transport Policies and the Decline of Bronzeville*, FINDING CHICAGO: GLOBAL PERSPECTIVES BLOG (Aug. 27, 2019), https://voices.uchicago.edu/findingchicago/2019/08/27/to-kill-a-neighborhood-urban-transport-policies-and-the-decline-of-bronzeville/#_ednref4 (explaining how federal policies favoring single-family suburban houses caused decline in funding for public transports, which in turn restricted the mobility of inner-city populations); Kalya Belsha, *Behind Sale of Closed Schools, a Legacy of Segregation*, THE CHICAGO REP. (Jan. 13, 2017), <https://www.chicagoreporter.com/behind-sale-of-closed-schools-a-legacy-of-segregation>; Alana Semuels, *Chicago's Awful Divide*, THE ATLANTIC (Mar. 28, 2018), <https://www.theatlantic.com/business/archive/2018/03/chicago-segregation-poverty/556649> (describing unprecedented school closures in Chicago's West Side).

143. See generally Douglas S. Massey & Jonathan Tannen, *A Research Note on Trends in Black Hypersegregation*, 52 DEMOGRAPHY 1025 (2015).

144. GREGORY ACS, ROLF PENDALL, MARK TRESKON & AMY KHARE, THE COST OF SEGREGATION: NATIONAL TRENDS AND THE CASE OF CHICAGO, 1990–2010 26–27 (2017),

gun violence and homicide rates, but the role of racial segregation in these problems cannot be disregarded or go unacknowledged. This is especially true when evaluating the efficacy of GIS technologies, like those used in the CPD pilot, because such assessments must also consider alternatives that were not explored or invested in and may have been better suited to address root cause issues or structural conditions that drive certain crimes.¹⁴⁵

For instance, research on racial segregation in Chicago determined that reducing the levels of black-White segregation could potentially reduce the city's homicide rate by thirty percent.¹⁴⁶ Notably, this reduction is similar to the crime reduction statistics cited by the CPD as attributable to the pilot, but the pilot did not change the citywide homicide rate, which, in fact, increased during the same observed period.¹⁴⁷ Thus, it is possible that some violent crimes were merely repositioned to other areas in or outside of the city that were not subjected to the same surveillance technologies and policing tactics as the neighborhoods in the CPD pilot. Although this inference is merely a conjecture, it demonstrates that the conclusions about the efficacy of GIS technologies made by CPD and researchers¹⁴⁸ may be specious and erroneous. Moreover, such positive yet fallacious evaluations of data-driven technologies further obscure how policies, practices, and tactics of policing serve to reinforce racial segregation and its consequences, just as described in Section III.A.1.¹⁴⁹

https://www.urban.org/sites/default/files/publication/89201/the_cost_of_segregation.pdf.

145. *See generally* JAMES FORMAN JR., LOCKING UP OUR OWN: CRIME AND PUNISHMENT IN BLACK AMERICA (2017) (detailing how law enforcement policy choices led to mass incarceration, specifically the decisions to increase criminalization instead of policies that aimed to expand social safety nets and community investment).

146. ACS ET AL, *supra* note 144, at 40.

147. Mclaughlin, *supra* note 131.

148. UNIV. OF CHICAGO URBANLABS, *Strategic Decision Support Centers (SDSCs)*, <https://urbanlabs.uchicago.edu/programs/strategic-decision-support-centers-sdscs> (“In 2017, Chicago experienced 764 fewer shooting incidents (22% reduction) relative to 2016. In District 007, historically one of the most violent districts in the city, these impressive gains are promising both compared to 2016 and historically . . .”) (last visited July 31, 2021); Ferguson, *supra* note 96, at 497. (“Early testing of HunchLab has shown positive results in Chicago and Philadelphia in reducing crime, but the findings have not been published in any peer-reviewed journals.”).

149. Roberts et al., *supra* note 97; *see also* Brief of Chicago Community-Based Organizations, Brighton Park Neighborhood Council, et al. as Amici Curiae Supporting Defendant at 25, *Illinois v. Williams*, 20 CR 0889601 (Cook Cnty. Cir. Ct. filed May 3, 2021), <https://endpolicesurveillance.com/documents/2021-05-03-Motion-for-Leave-to-File-Brief-as-Amici-Curiae-with-Ex.-A-Amicus-Brief-attached.pdf> (arguing for close judicial scrutiny of ShotSpotter's reliability because of its disproportionate impact on people of color in Chicago and contributory role in unconstitutional policing).

IV. CONCLUSION

This Article demonstrates a great need to critically examine the drivers of structural inequities and systemic disadvantage when evaluating issues related to technology and society. Unless we understand that the lack of diversity within technology and related sectors is an extension of broader societal patterns and problems, individuals who design, evaluate, and regulate technology and technology-mediated issues will continue to primarily come from dominant and privileged backgrounds. Similarly, if we continue to view algorithmic bias and other technology-mediated problems as technical issues, the proposed solutions will always be insufficient.

Data-driven technologies cannot be apolitical, ahistorical, or considered separate and distinct from social and power structures because technology, and scientific knowledge more generally, “embeds and is embedded in social practices, identities, norms, conventions, discourse, instruments and institutions”¹⁵⁰ Thus, interventions to improve the data-driven technology development process and to redress and prevent negative consequences of data-driven technologies must contend with the institutional and social practices that create unequal structural conditions and contribute to the differential treatment of individuals and groups, like racial segregation.

There are several emergent academic fields that seek to grapple with and mediate technological injustices (e.g., data justice and AI ethics, fairness, accountability, and transparency studies). But many scholars have questioned the adequacy of these fields and related discourse because they tend to examine discrimination, disadvantage, exclusion, misrecognition, hyper-surveillance, and other justice-related concerns primarily through technology.¹⁵¹ These fields and the interventions they produce generally have two issues. First, they fail to reckon with the disadvantages and harms that preceded and are often compounded by data-driven interventions. Second, they fail to decenter technology as the primary lens of analysis or modality of prevention and redress.¹⁵² Thus, much of this existing scholarship suffers from similar

150. Sheila Jasanoff, *The Idiom of co-production*, in STATES OF KNOWLEDGE: THE CO-PRODUCTION OF SCIENCE AND SOCIAL ORDER 1, 3 (Sheila Jasanoff ed., 2004).

151. See Matthew Le Bui & Safiya Umoja Noble, *We're Missing a Moral Framework of Justice in Artificial Intelligence: On the Limits, Failings, and Ethics of Fairness*, in THE OXFORD HANDBOOK OF ETHICS OF AI 164, 166–69, 176–77 (Markus D. Dubber, Frank Pasquale & Sunit Das eds., 2020); Seeta Peña Gangadharan & Jędrzej Niklas, *Decentering Technology in Discourse on Discrimination*, 22 INFO. COMM. & SOC'Y 882, 885–87 (2019); Anna Lauren Hoffman, *Where Fairness Fails: Data, Algorithms, and The Limits of Antidiscrimination Discourse*, 22 INFO. COMM. & SOC'Y 900, 903–09 (2019).

152. See Bui & Noble, *supra* note 151; Gangadharan & Niklas, *supra* note 151, at 882; Hoffman, *supra* note 151, at 900.

cognitive and epistemic gaps or biases detailed in this Article and risks entrenching structural inequality.

Within legal scholarship, there is a growing body of discourse that attempts to propose legal interventions that can provide a greater transparency, oversight, or mechanisms for contestation.¹⁵³ But most of these proposals are primarily procedural and fail to scrutinize the role of legal institutions, systems, policies, and practices in engendering the problems they seek to mitigate. As a result, much of this existing legal scholarship and the interventions it produces have epistemic, analytical, and practical issues. First, this scholarship ignores or fails to understand that many existing legal procedures and enforcement mechanisms are not suited to evaluate or effectively redress systemic harms or violations. And the underinclusive nature of many laws and regulations exist by design¹⁵⁴ or are the consequence of caselaw or other reforms that have diminished or distorted such laws' and regulations' utility.¹⁵⁵ These epistemic

153. See, e.g., Citron & Pasquale, *supra* note 76, at 18–30 (suggesting procedural safeguards for automated scoring systems); Kroll et al., *supra* note 77, at 696–99 (suggesting making algorithms reviewable and not simply complaint with specifications); Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 124–28 (2014) (arguing for procedural data due process, including notice, a hearing before an impartial adjudicator, and a judicial review, to mitigate predictive privacy harms caused by big data methodology); Danielle K. Citron & Ryan Calo, *The Automated Administrative State: A Crisis of Legitimacy*, 70 EMORY L.J. 797, 836–45 (2021) (suggesting an alternative approach to development of technology used by administrative agencies).

154. See, e.g., JILL STAUFFER, *ETHICAL LONELINESS: THE INJUSTICE OF NOT BEING HEARD*, 38–46 (2015) (describing how legal trials in general and adversarial systems in particular are often limited in their ability to evaluate and redress certain injustices or address certain harms because findings at trials are narrowly focused on perpetrators' legal culpability and do not create a full record of the underlying events); Frederik Zuiderveen Borgesius, *Strengthening Legal Protection Against Discrimination by Algorithms and Artificial Intelligence*, 24 INT'L J. HUM. RTS. 1572, 1576–85 (2020) (highlighting limitations in European non-discrimination and data-protection laws in addressing algorithmic bias and harms and proposing alternative regulations); Michael Veale & Frederik Zuiderveen Borgesius, *Demystifying the Draft EU Artificial Intelligence Act*, 22 COMPUTER L. REV. INT'L (Forthcoming 2021) (detailing the limitations and weaknesses of the European Commission's proposed Artificial Intelligence Act, including "cumulative" algorithmic harms that are not tied to one particular event but rather underlying patterns that will be difficult to prove under the proposed framework); Alan David Freeman, *Legitimizing Racial Discrimination Through Antidiscrimination Law: A Critical Review of Supreme Court Doctrine*, 62 MINN. L. REV. 1049, 1052–57 (1978) (arguing that antidiscrimination laws were conceived in "the perpetrator perspective" which has a myopic conception of racial discrimination and unnecessarily limits violations and remedies to instances where intent and causation are obvious and provable).

155. See, e.g., Charles R. Lawrence III, *The Id, the Ego, and Equal Protection: Reckoning with Unconscious Racism*, 39 STAN. L. REV. 317, 318–28 (1987) (arguing that constitutional law and its subsequent caselaw's focus on a blameworthy perpetrator means that many forms and incidents of discrimination will not be acknowledged or remedied through traditional legal recourse); Derrick Bell, *Racial Realism*, 24 CONN. L. REV. 363, 376–78 (1992) (arguing that

and analytical flaws in prevailing legal scholarship are consequential because many forms of algorithmic bias and harms are systemic, as demonstrated by the examples detailed in this Article. Moreover, several forms of algorithmic bias remain difficult to detect, discern, and prove,¹⁵⁶ rendering these harms invisible or beyond redress within a legal system that is more or less structured to permit or ignore some forms of discrimination and harm.¹⁵⁷ Second, the interventions proposed by this existing scholarship can only produce narrow remedies that allow some algorithmic and related harms to persist or only address some concerns related to algorithmic bias on a case-by-case basis.¹⁵⁸

racial progress towards equity is limited through traditional legal mechanisms because analysis of caselaw and real-world outcomes demonstrates that they effectively reinforce the status quo); Reva Siegel, *Why Equal Protection No Longer Protects: The Evolving Forms of Status-Enforcing State Action*, 49 STAN. L. REV. 1111, 1129–46 (1997) (detailing how legal reforms designed to defend protected statuses ultimately served to reinforce asymmetrical status relationships over time because social practices evolved to survive those legal reforms);

Privacy at The Margins, with Professor Scott Skinner-Thompson (Big Conversations), THE BTLJ PODCAST (Jan. 27, 2021), <https://btlj.org/2021/01/privacy-at-the-margins-with-professor-scott-skinner-thompson-big-conversations/> (detailing how privacy laws and theory have effectively diminished privacy of marginalized communities).

156. See, e.g., Emily Lane, *Mayor, Police Chief to Face Subpoenas from Convicted Gang Member over Palantir Claim*, THE TIMES-PICAYUNE, https://www.nola.com/news/crime_police/article_fa5949c4-a300-509d-90e8-2d7814f505f6.html (Jul. 12, 2019, 12:03 PM) (describing a defendant’s appeal of a gang-related conviction where the prosecutors allegedly withheld analytical evidence obtained through an AI system that was used in the police investigation and prosecution); Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove & Aaron Rieke, *Discrimination Through Optimization: How Facebook’s Ad Delivery Can Lead to Biased Outcomes*, 3 PROCS. ACM ON HUMAN-COMPUTER INTERACTION 199:1, 199:18–23 (2019) (finding that unknown mechanisms in Facebook’s ad delivery technology led to potentially discriminatory ad targeting outcomes even when advertisers were inclusive); Cyrus Farivar, *Tenant Screening Software Faces National Reckoning*, NBC NEWS (Mar. 14, 2021, 4:00AM), <https://www.nbcnews.com/tech/tech-news/tenant-screening-software-faces-national-reckoning-n1260975> (describing how tenant screening software improperly flagged a housing application as “unqualified” and did not provide an explanation for the classification to the property manager that rejected the tenant’s application).

157. See VIRGINIA EUBANKS, *AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR* 190–92 (2018) (describing how data-driven reverse redlining has replaced earlier forms of housing discrimination yet appears legally permissible since these tools do not explicitly use race to make decisions); Siegel, *supra* note 155, at 1113–48 (1997) (arguing that historical forms of discrimination and bias are often narrowly construed in caselaw, which exculpates present practices that entrench systems of social stratification); OSCAR H. GANDY, JR., *COMING TO TERMS WITH CHANCE: ENGAGING RATIONAL DISCRIMINATION AND CUMULATIVE DISADVANTAGE* 65–76 (2009) (describing several forms of discrimination that are legally and socially permissible).

158. Cf. RASHIDA RICHARDSON, JASON M. SCHULTZ & VINCENT M. SOUTHERLAND, *LITIGATING ALGORITHMS 2019 US REPORT: NEW CHALLENGES TO GOVERNMENT USE OF ALGORITHMIC DECISION SYSTEMS*, AI NOW INST. 8 (2019) (noting that a successful administrative due process challenge of a flawed algorithm used by Arkansas to determine and

This issue stems from the fact that procedural fixes to underinclusive legal frameworks will still fail to recognize and therefore remedy many forms of algorithmic harm.¹⁵⁹ And this problem is made worse by practical fragmentation of the American legal system (i.e., laws that can address algorithmic bias are dispersed amongst various statutes and regulatory authorities),¹⁶⁰ which can deter potential plaintiffs or insulate malefactors and complicit bystanders from accountability. Rather than addressing algorithmic bias, these proposed interventions also risk entrenching structural inequality while failing to reveal the full scope of remedial options available and needed by those harmed.

Instead, interventions and approaches should apply and add nuance to the acknowledgment and remediation of algorithmic bias and systemic disadvantage. They should also clarify and acknowledge that “technology assists and exists alongside, as opposed to at the center of a discriminatory and unjust society.”¹⁶¹ Yet, to achieve this we also need to change who is a part of and considered in data-driven technology development and the creation of interventions for technology-mediated problems. To advance these goals, the

allocate disability benefits did not provide relief to every injured benefit recipient); Press Release, Roderick & Solange MacArthur Justice Center, *As Lawsuit Over Chicago's Controversial Gang Database Comes to an End Organizations Turn to City Hall to Stop the Use of All CPD Gang Databases* (Sept. 3, 2020), <https://www.macarthurjustice.org/as-lawsuit-over-chicagos-controversial-gang-database-comes-to-an-end-organizations-turn-to-city-hall-to-stop-the-use-of-all-cpd-gang-databases/> (stating that a settlement of a lawsuit regarding the Chicago Police Department's gang databases was insufficient to redress algorithmic harms and will require continued advocacy to expose harms of the database and similar systems).

159. *E.g.*, Julie C. Suk, *Procedural Path Dependence: Discrimination and the Civil-Criminal Divide*, 85 WASH. U. L. REV. 1315, 1325 (2008) (arguing that procedural reforms fail to adequately redress discrimination because the existing legal system is not structured to adapt to changing social practices so “[o]ver time, some of these [discrimination] fact patterns may be a poor fit with the principles, policies, norms, and practices that pervade a particular procedural system”).

160. This fragmentation means that laws capable of addressing algorithmic bias are dispersed amongst various statutes and regulatory authorities. *See, e.g.*, Lydia X. Z. Brown, Michelle Richardson, Ridhi Shetty, Andrew Crawford & Timothy Hoagland, *Challenging The Use of Algorithm-Driven Decision-Making in Benefits Determinations Affecting People With Disabilities*, CTR. DEMOCRACY & TECH., 8–20 (2020) (highlighting various legal arguments and statutes that have been used to challenge harms resulting from the use of algorithm-driven decision-making in the public benefits context); Liz Richardson, *How FDA Regulates Artificial Intelligence in Medical Products*, PEW CHARITABLE TRUSTS (Aug. 5, 2021), <https://www.pewtrusts.org/en/research-and-analysis/issue-briefs/2021/08/how-fda-regulates-artificial-intelligence-in-medical-products> (describing the various ways the FDA regulates AI in healthcare and existing gaps in regulatory oversight where hospital accrediting bodies, standard-setting organizations, insurers, and other government agencies regulate patient safety).

161. Gangadharan & Niklas, *supra* note 151, at 886.

data-driven technology sector (i.e., artificial intelligence and automated decision-making) needs a transformative justice framework and praxis, which I will present and expound upon in future scholarship.

Transformative justice is a holistic approach and field of practice that seeks to address the root causes of harm and injustice and develop solutions that ultimately change social systems and structural conditions that contribute to or perpetuate harm and injustice writ large.¹⁶² Thus, in the context of the data-driven technologies discussed in this Article—predictive policing, school assignment algorithms, and GIS technologies—a transformative justice approach would not only seek technical or technological policy interventions but also seek harm reduction and remedial opportunities in criminal justice, education, urban planning, social welfare, and other areas. Transformative justice employs a systematic approach to problem analysis and incorporates some principles and practices of restorative justice, like the intentional inclusion of victims and other community members. But, unlike restorative justice, which seeks to restore the condition before a harm or injustice took place,¹⁶³ transformative justice requires a broader examination of collective responsibility in society for creating structural conditions and social practices that enable and perpetuate systemic harms and injustices.

By looking beyond the scope of state power as the primary response for redress, a transformative justice approach intentionally includes an assessment of individual and collective agency to both inflict harm and repair it. This broader inquiry can lead to a greater understanding of algorithmic harm, in addition to other relevant contexts, which can in turn shed more light on the complexities of the problem as well as opportunities for redress.¹⁶⁴ Unlike traditional legal approaches, where fact-finding and accountability are limited to an alleged perpetrator, a transformative justice approach can include and implicate complicit bystanders and institutions that should bear some

162. See generally ADRIENNE MAREE BROWN, *WE WILL NOT CANCEL US: AND OTHER DREAMS OF TRANSFORMATIVE JUSTICE* (2020) (examining the value of cancel culture and whether society needs to redress harm in a way that reflects its values); Anthony J. Nocella II, *An Overview of the History and Theory of Transformative Justice*, 6 *PEACE & CONFLICT REV.* (2011) (providing a historical, political, and philosophical overview of transformative justice); RUTH MORRIS, *STORIES OF TRANSFORMATIVE JUSTICE* (2000) (outlining the failures of the current penal system); PAUL GREADY & SIMON ROBINS, *FROM TRANSITIONAL TO TRANSFORMATIVE JUSTICE* (2019) (exploring transformative justice as an alternative to the frequently critiqued transitional justice doctrine).

163. See FANIA E. DAVIS, *THE LITTLE BOOK OF RACE AND RESTORATIVE JUSTICE: BLACK LIVES, HEALING, AND US SOCIAL TRANSFORMATION* 19–29 (2019).

164. Cf. STAUFFER, *supra* note 154, at 45–49 (highlighting how the broader fact-finding enabled through truth commissions can build a more comprehensive narrative of harms and identify more sites where repair is need).

responsibility in reparative interventions as well as be fully accountable for what is needed to redress the compounded and systemic harms related to algorithmic bias.¹⁶⁵ Such comprehensive and shrewd analysis is necessary because many algorithmic harms stem from or are related to long-standing, systemic issues (e.g., racial segregation, poverty, and police misconduct) that are not only the result of bad or misguided actors' behavior but also of bystanders' facilitation or ignorance of harms, as demonstrated in this Article. Therefore, a transformative justice approach can facilitate radical social changes, as well as a variety of technical and non-technical interventions that can adequately confront the intersectional and intergenerational nature of technology-mediated problems and withstand the current pace of innovation.

Though this proposal may seem idealistic or even heterodox for technology development and regulation, it is necessary in light of who dominates all relevant sectors driving technology development and technology policy interventions, in addition to the epistemic gaps, highlighted in this Article and other critical scholarship,¹⁶⁶ that hinder the formation of meaningful solutions. A transformative justice framework and praxis have three advantages. First, they can force visibility of power structures and dynamics that are often opaque yet stymie necessary reforms or actions. Second, the transformative justice framework and praxis can center people and perspectives that are typically excluded from but pivotal to the problem formulation process of data-driven technology development. Third, the transformative justice framework and praxis can lead to systemic solutions that not only address technical concerns but underlying root causes that sustain the status quo. Therefore, in the context of data-driven technology development and policy, the transformative justice framework and praxis can help us advance towards a future where technology and society are designed for collective belonging.

165. *Cf. id.* at 34–68 (describing limitations of legal trials and truth commissions for addressing long-standing or systemic injustices).

166. *See* sources cited *supra* note 151.

ALLOCATING RESPONSIBILITY IN CONTENT MODERATION: A FUNCTIONAL FRAMEWORK

Deirdre K. Mulligan[†] & *Kenneth A. Bamberger*^{††}

ABSTRACT

This Article develops a framework for both assessing and designing content moderation systems consistent with public values. It argues that moderation should not be understood as a single function, but as a set of subfunctions common to all content governance regimes. By identifying the particular values implicated by each of these subfunctions, it explores the appropriate ways the constituent tasks might best be allocated—specifically to which actors (public or private, human or technological) they might be assigned, and what constraints or processes might be required in their performance. This analysis can facilitate the evaluation and design of content moderation systems to ensure the capacity and competencies necessary for legitimate, distributed systems of content governance.

Through a combination of methods, legal schemes delegate at least a portion of the responsibility for governing online expression to private actors. Sometimes, statutory schemes assign regulatory tasks explicitly. In others, this delegation often occurs implicitly, with little guidance as to how the treatment of content should be structured. In the law’s shadow, online platforms are largely given free rein to configure the governance of expression.

Legal scholarship has surfaced important concerns about the private sector’s role in content governance. In response, private platforms engaged in content moderation have adopted structures that mimic public governance forms. Yet, we largely lack the means to

DOI: <https://doi.org/10.15779/Z383B5W872>

© 2021 Deirdre K. Mulligan & Kenneth A. Bamberger.

† Professor, School of Information, UC Berkeley; Faculty Co-Director, Berkeley Center for Law and Technology, Berkeley, School of Law; Co-Director, Algorithmic Fairness and Opacity Group, School of Information, UC Berkeley. Many thanks to the following persons for thoughtful feedback: organizers David Kaye and Gregory Shaffer, discussant KS Park, and participants at the University of California, Irvine, School of Law Symposium on the Transnational Legal Ordering of Privacy and Speech; moderator David Vladeck and co-panelists Dina Srinivasan and Ashkan Soltani at the National Academies of Sciences, Engineering and Medicine’s Committee on Science, Technology, and Law workshop “Section 230 Protections: Can Legal Revisions or Novel Technologies Limit Online Misinformation and Abuse?”; organizers and participants at the 25th Annual BCLT/BTLJ Symposium—Lex Informatica: The Formulation of Information Policy Rules through Technology. Special thanks to Daphne Keller, Nicole Wong, Joris van Hoboken, and Naomi Appelman for providing feedback on subsequent drafts, and to Jessica Li, Khash Goshtasbi, and Sophia Wallach for their detailed feedback, editing, and patience. Research for this article, and the development of the handoff model has been funded by generous support from the US NSF INSPIRE SES1537324.

†† The Rosalinde and Arthur Gilbert Foundation Professor of Law, University of California, Berkeley; Faculty Co-Director, Berkeley Center for Law and Technology.

measure whether these forms are substantive, effectively infusing public values into the content moderation process, or merely symbolic artifice designed to deflect much needed public scrutiny.

This Article’s proposed framework addresses that gap in two ways. First, the framework considers together all manner of legal regimes that induce platforms to engage in the function of content moderation. Second, it focuses on the shared set of specific tasks, or subfunctions, involved in the content moderation function across these regimes.

Examining a broad range of content moderation regimes together highlights the existence of distinct common tasks and decision points that together constitute the content moderation function. Focusing on this shared set of subfunctions highlights the different values implicated by each and the way they can be “handed off” to human and technical actors to perform in different ways with varying normative and political implications.

This Article identifies four key content moderation subfunctions: (1) definition of policies, (2) identification of potentially covered content, (3) application of policies to specific cases, and (4) resolution of those cases.

Using these four subfunctions supports a rigorous analysis of how to leverage the capacities and competencies of government and private parties throughout the content moderation process. Such attention also highlights how the exercise of that power can be constrained—either by requiring the use of particular decision-making processes or through limits on the use of automation—in ways that further address normative concerns.

Dissecting the allocation of subfunctions in various content moderation regimes reveals the distinct ethical and political questions that arise in alternate configurations. Specifically, it offers a way to think about four key questions: (1) what values are most at issue regarding each subfunction; (2) which activities might be more appropriate to delegate to particular public or private actors; (3) which constraints must be attached to the delegation of each subfunction; and (4) where can investments in shared content moderation infrastructures support relevant values? The functional framework thus provides a means for both evaluating the symbolic legal forms that firms have constructed in service of content moderation and for designing processes that better reflect public values.

TABLE OF CONTENTS

I.	INTRODUCTION	1093
II.	SHIFTING THE FOCUS FROM FORM TO FUNCTION.....	1102
A.	LEGALISTIC FORM AS A RESPONSE TO LEGITIMACY CRITIQUES ...	1102
1.	<i>Google’s Advisory Council on the Right to be Forgotten</i>	1103
2.	<i>Facebook’s Oversight Board</i>	1107
3.	<i>Transparency Reports</i>	1110
B.	SHIFTING THE FOCUS FROM FORM TO FUNCTION	1114
III.	DEVELOPING A FUNCTIONAL FRAMEWORK.....	1117
A.	GROUNDING A FUNCTIONAL APPROACH.....	1118
1.	<i>Theoretical Roots.....</i>	1118
2.	<i>Descriptive Realities</i>	1120
B.	THE FUNCTIONAL FRAMEWORK	1121
1.	<i>Components of A Functional Framework.....</i>	1122
a)	Identifying the Subfunctions of Content Moderation..	1122
b)	Examining the Values Implicated, and Competencies Required, by the Different Subfunctions	1125

c)	Constraints Intended to Protect Values and Enhance Competencies.....	1127
2.	<i>Understanding Elements of the Functional Framework Through Case Studies.....</i>	1127
a)	Case Studies of Subfunctions	1128
i.	<i>Defining: Section 230 and the DMCA.....</i>	1128
b)	Identifying: RTBF and CSAM	1130
c)	Applying: Section 230 and the DMCA	1137
d)	Resolving: § 230 and the DMCA.....	1140
C.	LESSONS FROM THE CASE STUDIES: THE TYPES OF CONSTRAINTS USED IN STRUCTURING SUBFUNCTIONS.....	1141
a)	Process Constraints.....	1141
b)	Constraints on the Allocation of Functions Between Particular Technical and Human Actors	1143
IV.	APPLYING THE FUNCTIONAL FRAMEWORK	1148
A.	EVALUATING THE SYMBOLIC STRUCTURES	1148
1.	<i>Google Advisory Council.....</i>	1150
2.	<i>Facebook Oversight Board.....</i>	1154
3.	<i>Transparency Reports.....</i>	1159
B.	THE CONSTRUCTIVE TURN: USING A FUNCTIONAL FRAMEWORK TO CONFIGURE CONTENT MODERATION.....	1164
1.	<i>Leveraging Competencies and Addressing Democratic Deficits: Reimagining the Global Internet Forum to Counter Terrorism.....</i>	1165
V.	CONCLUSION.....	1170

I. INTRODUCTION

Addressing harms from online content has proven to be a major puzzle of the digital age. It confounds traditional notions of regulation, as First Amendment limits and other non-intervention norms combine with the trans-jurisdictional scope of platform operations to circumscribe the public role in governance of platform content. Through a combination of methods, legal schemes delegate at least a portion of the responsibility for governing expression to private actors. Sometimes, statutory schemes assign regulatory tasks explicitly. Yet this delegation often occurs implicitly, with little guidance as to how the treatment of content should be structured. For example, § 230 of the Communications Decency Act¹ empowers platforms to moderate content by shielding them from liability without specifying processes for, or

1. 47 U.S.C. § 230.

even the content subject to, moderation. In the law's shadow, online platforms are largely given free reign to configure the governance of expression.

Legal scholarship has surfaced important concerns about the private sector's role in content governance. This work includes critiques of particular company processes as opaque, arbitrary, or untethered to important public norms regarding the governance of expression.² It also criticizes the use of technology to automate content moderation—a particular form of privatization—for its opacity and for the overbroad, discriminatory, and arbitrary outcomes it produces.³

In response to critiques, private platforms engaged in content moderation have adopted structures that mimic public governance forms. One example, Google's Right to be Forgotten Advisory Council involved external experts in the development of rules and processes to guide the private content

2. evelyn douek, *The Limits of International Law in Content Moderation*, 6 U. CALIF. IRVINE J. INT'L., TRANSNATIONAL, & COMPARATIVE L. 37 (2021) (“[I]nwards-looking, largely public relations-oriented content governance models so widely deployed today’ . . . are largely untethered from any particular normative commitments, leaving them unconstrained and arbitrary.”).

3. Emma Llansó, Joris van Hoboken, Paddy Leerssen & Jaron Harambam, *Artificial Intelligence, Content Moderation, and Freedom of Expression* (Feb. 26, 2020) (Working Paper, Transatlantic Working Grp., Bellagio Session), <https://www.ivir.nl/publicaties/download/AI-Llanso-Van-Hoboken-Feb-2020.pdf>; see Emma J. Llansó, *No Amount of “AI” in Content Moderation Will Solve Filtering’s Prior-Restraint Problem*, 7 BIG DATA & SOC’Y 1 (2020) (arguing that the use of technology to proactively moderate content “acts as a prior restraint on speech, regardless of the accuracy”); Robert Gorwa, Reuben Binns & Christian Katzenbach, *Algorithmic Content Moderation: Technical And Political Challenges In The Automation Of Platform Governance*, 7 BIG DATA & SOC’Y 1 (2020) (arguing algorithmic content moderation increases the opacity of platform practices, exacerbates concerns with fairness and accountability, and obscures important political choices); Hannah Bloch-Wehba, *Automation in Moderation*, 53 CORNELL INT’L L.J. 41 (2020) (arguing ex-ante algorithmic moderation expands platforms’ unaccountable control over online speech, exacerbating risks to freedom of speech and association, privacy, and equality). This important literature has produced a wealth of information about the operation of specific instantiations of the content moderation function. It has, moreover, identified and framed important questions implicated by content moderation generally, including the risks of privatizing functions traditionally performed by government entities, the substantive and procedural deficits of platform moderation practices, and the hazards posed by various technical methods of natural language processing used to automate content moderation, such as image recognition and cryptographic hashing. Common themes across this work are concerns with the lack of consistent commitment to normative touchstones, such as international human rights norms, and the lack of attention to context—be it the histories of violence or oppression within particular communities, the events and conversations in which platform content is mobilized, or the divergent meaning and importance of content over time and between communities.

moderation operation.⁴ A second example, the most recent and grandiose such structure, is the Facebook Oversight Board (FBOB), frequently referred to as its “Supreme Court.”⁵ A third example, the most widespread, are “transparency reports,” consisting of company-issued public reports disclosing content moderation outcomes along with data about other firm practices that affect the privacy and freedom of expression interests of users.⁶

Longstanding research in “new governance” points to the persistent role of private firms and the need to think constructively, rather than reactively, about the assets and deficits private actors bring to discrete aspects of authority.⁷ At the same time, socio-legal scholars warn that governance in private hands can produce symbolic or ceremonial structures that imbue corporate acts with apparent legitimacy but do little to further the public values at stake.⁸ Criticism of transparency reporting and the FBOB resonates with these insights.⁹

4. Google convened the Council in 2015 in response to the European Union’s ruling in *Google Spain v. Agencia Española de Protección de Datos*. Case C131 /12 Google Spain v. Agencia Española de Protección de Datos (AEPD), ECLI:EU:C:2014:317 (May 13, 2014). The Council is no longer operational. See Carol A. F. Umhoefer, *Europe: Right to be Forgotten—Google Advisory Council published its report*, Lexology.com (available at <https://www.lexology.com/library/detail.aspx?g=db11a725-6a65-4250-b792-9a93d4394057>) (noting that the Advisory Council’s Final Report was published on February 6, 2015).

5. Facebook convened the Facebook Oversight Board in 2020, and the Board is still in operation today. See OVERSIGHT BOARD, <https://oversightboard.com/> (last visited Apr. 19, 2022).

6. See, e.g., Robert Gorwa & Timothy Garton Ash, *Democratic Transparency in the Platform Society*, in *SOCIAL MEDIA AND DEMOCRACY: THE STATE OF THE FIELD AND PROSPECTS FOR REFORM* 286, 293–299 (Nate Persily & Josh A. Tucker eds., 2020) (discussing major voluntary transparency initiatives of platform companies); Camille François & evelyn douek, *The Accidental Origins, Underappreciated Limits, and Enduring Promises of Platform Transparency Reporting About Information Operations*, 1 J. ONLINE TR. & SAFETY 1, 4–11 (2021) (discussing the history and development of transparency disclosures regarding platforms’ actions to address “information operations,” an ambiguous category used by platforms to address coordinated, deceptive activity typically targeting clusters of accounts). The three major platforms discussed in this article publish transparency reports to provide the public with some information about content removals, among other actions. See *infra* Section III.A.3.

7. See *infra* note 20.

8. See Lauren B. Edelman, *Legal Ambiguity and Symbolic Structures: Organizational Mediation of Civil Rights Law*, 97 AM J. SOC. 1531, 1542 (1992) (exploring the ways that “[l]aws that are ambiguous, procedural in emphasis, and difficult to enforce invite symbolic responses—responses designed to create a visible commitment to law, which may, but do not necessarily,” further public values).

9. See, e.g., Monika Zalnieriute, “Transparency-Washing” In *The Digital Age: A Corporate Agenda of Procedural Fetishism*, 8 CRITICAL ANALYSIS L. 39 (2021) (critiquing Transparency Reports of IBM, Google, and Facebook as “procedural fetishism” that provides limited substantive protection).

Rigorous assessment of the “effectiveness” of the structures adopted by platforms requires clarity about the deficits they are meant to address in the private content moderation process. Yet it is often unclear, at a relevant level of specificity, exactly which deficits new structures are targeted to address and against which yardsticks private content moderation systems should be measured. The lack of a clear metric for evaluation stymies efforts to develop and evaluate alternative approaches to structuring and constraining private content moderation and to envision the appropriate role for public law in mandating or catalyzing those alternatives.¹⁰

This Article suggests two analytic shifts that enable the assessment of content moderation systems’ alignment with public values and that enable the design of content moderation systems that ensure the capacity and competencies necessary for legitimate, distributed systems of content governance. Both shifts draw attention to content moderation’s function, rather than its form.¹¹ The first shift involves zooming out—that is,

10. International human rights law has become the dominant framework for assessing platform content governance. See, e.g., David Kaye, *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, No. A/HRC/38/35 (June 2018) ¶ 41 (arguing for the implementation of “human rights standards transparently and consistently, with meaningful user and civil society input” as a means for holding “both States and companies accountable to users across national borders”); Molly K. Land, *Against Privatized Censorship: Proposals for Responsible Delegation*, 60 VA. J. INT’L L. 363 (2020) (finding that many regulatory regimes that deputize platforms to police and govern content are unlawful under human rights law and proposing a “human rights non-delegation doctrine”); Barrie Sander, *Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation*, 43 FORDHAM INT’L L.J. 939, 966–68 (2020); Evelyn Aswad, *The Future of Freedom of Expression Online*, 17 DUKE L. & TECH. REV. 26, 57–67 (2018). For a discussion of the limits of international human rights law as a tool for reforming content moderation, see douek, *supra* note 2, at 50–64 and Sander, *supra* note 10, at 968–70. Hannah Bloch-Wehba argues that platforms should adopt basic principles of administrative law to ensure accountability to the public. *Global Platform Governance: Private Power in the Shadow of the State*, 72 SMU L. REV. 27 (2019). evelyn douek further notes that “requirements of a clear, precise, and transparent statement of a rule that is justified in the pursuance of a legitimate purpose . . . and due process requirements” are not unique to international human rights or administrative law. douek, *supra* note 2, at 63.

11. Recent work by Niva Elkin-Koren and Maayan Perel points to a different way a focus on functions in the regulation of online expression might be important. See *Separation of Functions for AI: Restraining Speech Regulation by Online Platforms*, 24 LEWIS & CLARK L. REV. 857 (2020). In particular, they recommend distinguishing between “public” functions related to the identification and removal of unlawful content from “private” functions driven by business-related content moderation imperatives, both of which rely on the use of artificial intelligence. These authors suggest that the former should only utilize “independent” AI tools that “embed[] public policy.” *Id.* at 857–58. In a similar vein, Barrie Sander grounds his human rights analysis by first delineating four platform content moderation activities in which

considering together all manner of legal frameworks that induce platforms to moderate online content. The second involves zooming in—that is, focusing on the shared set of specific tasks, or subfunctions, involved in the content moderation function across these regimes.

Our first analytic shift—zooming out by considering together any legal regimes that structures the governance of online expression—focuses on the functional output rather than the doctrinal basis or problem statement of content moderation. Online content moderation regimes intended to reduce harm arise in a range of contexts. Some, such as those arising under § 230, are widely recognized and labeled as “content moderation” regimes. Others, although they too govern online content, have largely escaped such identification and are generally treated as components of other legal subjects including copyright law, privacy law, human trafficking and child exploitation law, terrorism law, and harassment law.¹²

The wide range of existing statutory schemes that drive the governance of content reflect an extensive variety of approaches.¹³ Legal frameworks vary by specificity and tactic—from the detailed regulatory scheme set out in the Digital Millennium Copyright Act (DMCA)¹⁴ to the General Data Protection Regulation’s (GDPR) less-specified use of liability risks pursuant to its Right to be Forgotten mandate.¹⁵ Some frameworks directly mandate *what* content should be moderated, such as terrorism-related material, child sexual abuse material,¹⁶ and hate speech. Some specify explicitly *who* should be involved,

transparency and oversight should be provided: rulemaking, decision-making, content and advertising, and regulatory compliance. See Sander, *supra* note 10 at 998–90.

12. A vast array of other laws can be the source of content removal requests. For example, they can be U.S. Security and Exchange Commission regulations, trade secret law, fraud, or false advertising laws as well as speech that aids or abets illegal conduct (instructional content for criminal or negligent activity).

13. For one taxonomy of approaches states use to enlist private actors in regulating online content, see Molly K. Land, *Against privatized censorship: proposals for responsible delegation*, 60 VA. J. INT’L L. 363, 399 (2019) (“[States use] command and control, intermediary liability, and extra-legal influence.”)

14. Digital Millennium Copyright Act (DMCA), 17 U.S.C. §§ 512, 1201–05, 1301–32.

15. Regulation (EU) 2016/679, of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), art. 17, 2016 O.J. (L 119) 1, 43 (detailing the “Right to erasure (‘right to be forgotten’)” by providing that “[t]he data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay and the controller shall have the obligation to erase personal data without undue delay” where one several grounds applies).

16. Although the term child pornography is still used in legislation in the United States and elsewhere, for our discussion we use the term Child Sexual Abuse Material (CSAM) to

allocating discrete tasks within the content moderation process to a range of public and private actors (i.e., NGOs, content creators, courts, users, and other individuals). Some mandate particular procedural constraints regarding *how* the content moderation process should proceed.

Broadening the lens to include a spectrum of varied content moderation regimes surfaces differences between those regimes. Such analysis highlights the range of approaches to content moderation and offers examples that can be compared to improve content moderation's function.

Yet examining a broader range of content moderation regimes together also points to similarities amongst the different regimes. Specifically, it highlights the existence of distinct common tasks and decision points that together constitute the content moderation function across contexts—tasks and decision points that can each be assigned to different actors and structured differently.

Our second analytic shift—zooming in—focuses on the shared set of tasks, or “subfunctions,” involved in the content moderation function across regimes. Content moderation is not a single undifferentiated function. It is comprised of a variety of subfunctions—individual tasks and decisions that may be “handed off” to different human and technical actors to perform in different ways with different normative and political implications. This shift towards a focus on the allocation of responsibility constitutes an application of the “handoff model” for evaluating sociotechnical systems in ethical and political terms developed by Dierdre Mulligan, one of this Article's authors, and information scholar Helen Nissenbaum.¹⁷

The handoff model focuses on identifying the constituent tasks that comprise systems, here those involving content moderation and the different values implicated by each task. It then considers four questions: What values are important to preserve in structuring those tasks or decisions? Which actors (human or technical), then, should be given the power to perform the task or decision? What procedural or technical constraints should structure the performance of the task or decision? Do the constraints protect or promote the public norms at issue?

This Article uses the handoff model to identify and focus on four key content moderation subfunctions: (1) *definition* of policies, (2) *identification* of potentially covered content, (3) *application* of policies to specific cases, and (4)

acknowledge that this material is distinct from adult pornography and depicts child abuse and exploitation.

17. Deirdre K. Mulligan & Helen Nissenbaum, *The Concept of Handoff as a Model for Ethical Analysis and Design*, in THE OXFORD HANDBOOK OF ETHICS OF AI 233 (Markus D. Dubber, Frank Pasquale & Sunit Das eds., 2020) [hereinafter *The Concept of Handoff*].

resolution of those cases. This list does not represent an exclusive typology of the stages of content moderation; others might categorize and label subfunctions differently. But these subfunctions reflect the structure of many of the statutes that drive content moderation and reflect insights from the authors' participation in policy processes around the adoption of key legal frameworks driving private platforms' content moderation practices.

Focusing on subfunctions illuminates the ways that the allocation of responsibility to different public or private actors to perform discrete subfunctions can improve legitimacy. Using these four subfunctions supports a rigorous analysis of how to leverage the capacities and competencies of government and private parties throughout the content moderation process. Such attention also highlights how the exercise of that power can be constrained—either by requiring the use of decision-making processes or through limits on the use of automation—in ways that further address normative concerns.

Identifying different subfunctions surfaces three things. First, it illuminates possible choices regarding the allocation of tasks to different actors. As an example, the DMCA assigns the task of identification of potentially infringing material online to content creators.¹⁸ The regime developed to address Child Sexual Abuse Material (CSAM), by contrast, in part relies on machine-learning systems for this subfunction.¹⁹

Second, focusing on subfunctions makes visible the normative implications of different content moderation configurations. Each identified subfunction implicates different governance norms and demands a different set of competencies. Thus, an assessment of any content moderation regime requires an inquiry into the appropriateness of assigning each subfunction. (For example, which assigned actor has the appropriate incentives, relevant information, or technical capacity in any given context?)

Finally, focusing on discrete subfunctions surfaces constraints, which may take the form of required procedures or limits on automation that can limit different actors' power to perform those tasks. For example, although it may be appropriate to allow a platform to rely on users or a machine-learning system to identify material for review under a content moderation policy, allowing reliance on either to define the content subject to moderation may raise concerns.

Thus, this Article's functional framework generates constructive insights that are both concrete and generalizable across contexts. It offers a means for

18. See 17 U.S.C. § 512 (establishing the notice and takedown procedure).

19. See *infra* text accompanying notes 130–139.

critically assessing the way various content moderation regimes both allocate and constrain various subfunctions. It identifies the implications of those different arrangements for public values. And it considers how these subfunctions might be appropriately structured to forge legitimate content governance systems going forward. This Article proceeds in the following way.

Part I explores three structures that private platforms have adopted in response to legitimacy critiques: Google’s Advisory Council on the Right to be Forgotten, Facebook’s Oversight Board, and the use of transparency reports. Noting the shortcomings of existing frameworks to assess the effectiveness and appropriateness of these regimes, this Part establishes the need for an evaluative framework that focuses on the subfunction deficits of the content moderation process that each structure attempts to address.

Part II sets forth this new evaluative framework. It first identifies and discusses four key subfunctions, the public values each implicates, and the competencies each requires. It then explores case studies from diverse content moderation regimes to illustrate possible task allocations to ensure that the appropriate competencies are brought to bear. Examples of constraints that in different circumstances have been—or could be—imposed on the performance of each subfunction include limits both on the process and on the identity of the actors (human and technical) by whom subfunctions may be implemented. Thus, the framework considers ways that largely private decision systems can best be held accountable to key normative values.

This framework reflects the insights of our and others’ work in “new governance” scholarship, emphasizing the importance of identifying, surfacing, structuring, and constraining decision-making delegated to private actors in governance in light of accountability to public norms.²⁰ By engaging in a precise exploration of how each actor enlisted in a content moderation regime executes the function delegated to it, the framework allows consideration of the ways in which replacing one actor with another disrupts (or doesn’t disrupt) the ethical and political dimensions of the subfunction and the configuration of values in the system as a whole.

20. See, e.g., Kenneth A. Bamberger, *Technologies of Compliance: Risk and Regulation in a Digital Age*, 88 TEX. L. REV. 669, 684 (2010) [hereinafter Bamberger, *Technologies of Compliance*] (focusing on the “decisionmaking processes of private actors” acting “as partners in regulation”); Kenneth A. Bamberger, *Regulation as Delegation: Private Firms, Decisionmaking, and Accountability in the Administrative State*, 56 DUKE L.J. 377 *passim* (2006) [hereinafter Bamberger, *Regulation as Delegation*]; Kenneth A. Bamberger & Deirdre K. Mulligan, *New Governance, Chief Privacy Officers, and the Corporate Management of Information Privacy in the United States: An Initial Inquiry*, 33 L. & POL’Y 477, 480–82 (2011) (summarizing the “new governance” literature).

Finally, Part III applies this analytic framework. First, it evaluates to what extent (if at all) the symbolic content moderation forms discussed in Part II provide competencies that address the relevant democratic deficiencies in the subfunction at issue. Second, making a constructive turn, it uses the Global Internet Forum to Counter Terrorism Shared Industry Hash Database²¹ to illustrate how to allocate subfunctions and coordinate subfunction constraints. As an example, it applies the framework to the Global Internet Forum to Counter Terrorism Shared Industry Hash Database. It then compares that content moderation configuration to the National Center for Missing and Exploited Children's Hash Database to highlight the different characteristics of regulated content and the different allocations of responsibility for the databases. This comparison illustrates the ways that the choice about which actor should be given authority for those databases affects the extent to which investments in shared content moderation infrastructures can support public values, including transparency, legitimacy, nondiscrimination, rational decision-making, and the promotion of competition.

Dissecting the allocation of subfunctions in various content moderation regimes reveals the distinct ethical and political questions that arise in alternate configurations. Specifically, it offers a way to think about four key questions: (1) what values are most at issue regarding each subfunction; (2) which activities might be more appropriate to delegate to particular public or private actors; (3) which constraints need to attach to the delegation of each subfunctions; and (4) where investments in shared content moderation infrastructures could support relevant values? The functional framework thus provides a means for evaluating the symbolic legal forms that firms have constructed in service of content moderation.

This Article's functional framework applies the handoff model to the content moderation landscape. Too often the salient differences in a new or competing functional arrangement of content moderation comes to light only after adoption. A functional framework offers a means to frontload this values analysis, allowing regulators, system designers, and other stakeholders to foresee, at least to some extent, and prioritize values during design. Looking forward, this Article's proposed framework enables analyses of how the mix of actors used to perform a function matters even before a content moderation regime is put into place.

21. *Tech Innovation*, GLOB. INTERNET F. TO COUNTER TERRORISM, <https://gifct.org/tech-innovation> (describing the hash-sharing database).

II. SHIFTING THE FOCUS FROM FORM TO FUNCTION

This Part first describes the structures platforms have adopted to insulate their private content moderation practices from critiques claiming they fail to sufficiently satisfy public governance values. It next unpacks the specific content moderation *subfunction* that each structure seeks to legitimate and, with this insight, suggests a new analytic direction for evaluating the alignment of content moderation systems with public values.

A. LEGALISTIC FORM AS A RESPONSE TO LEGITIMACY CRITIQUES

Faced with critiques questioning the legitimacy of their private content moderation activities,²² social media platforms have attempted to reshape the roles and responsibilities for moderating content online by evolving what social and technical studies (STS) theorists would term content management “scripts.”²³ Most visibly, these scripts include what legal sociologist Lauren Edelman refers to as “symbolic legal structures”—organizational approaches that evoke a notion of legality.²⁴ Through the adoption of structures that mimic those of public legal institutions—such as reports, public hearings, and appellate review structures—platforms seek to convey a sense of legitimacy to actions that are frequently entirely unrelated to public legal mandates.

This Section describes three paradigmatic examples of symbolic legal structures:

- (1) Google’s Advisory Council on the Right to be Forgotten (GAC)
- (2) Facebook’s Oversight Board (FBOB)
- (3) Transparency Reports (TR)²⁵

These structures mimic legal institutions and claim attributes often associated with public governance: government advisory committees/expertise and stakeholder participation (GAC); courts/expertise and independence (FBOB); and public access to the outcomes of adversary

22. *See supra* notes 2–3.

23. MADELEINE AKRICH, THE DE-SCRIPTION OF TECHNICAL OBJECTS, *in* SHAPING TECHNOLOGY/BUILDING SOCIETY: STUDIES IN SOCIOTECHNICAL CHANGE 205 (Wiebe E. Bijker & John Law eds., 1992).

24. LAUREN B. EDELMAN, WORKING LAW: COURTS, CORPORATIONS, AND SYMBOLIC CIVIL RIGHTS 101–02 (2016).

25. Platforms use Transparency Reports to publicly disclose data about the outcomes and legal bases of content removal requests they receive from different jurisdictions around the world, and, in a few instances, some information about removals pursuant to platform policies.

processes, particularly those where the government is a party/transparency (TR).

1. *Google's Advisory Council on the Right to be Forgotten*

Google's Advisory Council on the Right to be Forgotten arose in response to the European Court of Justice's 2014 decision in *Google Spain SL v. Agencia Española de Protección de Datos* (hereinafter *Google Spain*).²⁶ The ruling clarified that Google, as a data controller under data protection law, had an independent obligation to comply with the General Data Protection Regulation, Europe's data protection law. The court reasoned that because the information was no longer necessary for the purpose of real-estate auction, the plaintiff's interest in privacy overrode the "interest of the general public" in having access to private information. Therefore, the information had to be removed from Google's search results.

The holding left Google in the difficult position of having to develop decision criteria and processes for handling privacy objections to search engine results. Specifically, the ruling required Google to evaluate going forward whether an individual has a cognizable data protection interest and weigh that interest against "the preponderant interest of the general public in having . . . access to the information."²⁷ Thus Google was thrust into definitional work—policy formation—not just implementation.

The company objected to the role the court asked it to play throughout the *Google Spain* dispute, and many advocates and scholars shared Google's concerns. However, the concerns about the assignment to private actors of the authority to engage in such definitional work are not unique to the right to be forgotten.

26. Case C131 /12 *Google Spain v. Agencia Española de Protección de Datos* (AEPD) (*Google Spain*), ECLI:EU:C:2014:317 (May 13, 2014), <http://curia.europa.eu/juris/document/document.jsf?text=&docid=152065&doclang=EN> (finding Google had an obligation to delist search results that would normally be returned in response to queries on an individual's name where those results interfered with the privacy of the individual). *Id.* The E.U. Data Protection Directive developed guidance through various working groups to help companies comply with the European General Data Protection Regulation (GDPR). The GDPR, which came into force on May 25, 2018, adopted the right to be forgotten framework developed in the European Court of Justice's decision. 2016 O.J. (L 119) 1.

27. While this is often discussed as a privacy interest, it is more specifically a data protection interest arising under art. 8 rather than art. 7 of the Charter of Fundamental Rights of the European Union, 2012 O.J. (C 326) 2. For a useful discussion of the distinction and its implications for content removals by online service providers, see Daphne Keller, *The Right Tools: Europe's Intermediary Liability Laws and the EU 2016 General Data Protection Regulation*, 33 BERKELEY TECH. L. J. 287, 315–18 (2018).

Under many other legal frameworks, platforms are responsible for determining which content is subject to removal under the relevant policies (public or private). Liability regimes that put platforms at risk but give them no, or at least limited, guidance about how to apply broad legal standards delegate significant definitional work to platforms, like the *Google Spain* decision that led to the creation of the GAC. Moreover, legal frameworks like Section 230, which shield content moderation activities of social media platforms irrespective of the policies or processes they use to do it, delegate nearly all definitional work to platforms.

Faced with a regulatory system that delegates responsibility for developing both substantive decision-making criteria and processes for enforcing public law (rather than platform policy), Google employed a structure that resembled those used by government entities to develop policy: the government advisory committee. Google appointed an advisory council to design a system to process removal requests and to provide substantive decision-making criteria regarding how to balance an individual's right to privacy with the public's interest in access to information in the moderation of content implicated by the right to be forgotten.²⁸

In so doing, the company chose to adopt a form that resembled a common symbolic legal structure that government bodies use to garner advice about complex and often contentious policy choices. From federal advisory committees to congressionally created commissions, government entities frequently enlist outside experts in the assessment of substantive policy matters to ensure that expert knowledge and the perspectives of multiple constituencies are included and that the committee's work is relatively transparent and deliberative. For example, the Federal Advisory Committee Act (FACA),²⁹ passed in 1972, both (1) provides for the involvement of committees composed of experts, representatives of stakeholders, and representatives with different political views as tools for providing policy advice to U.S. government entities, and (2) imposes constraints on the composition of committees requiring them to be "fairly balanced in terms of the points of view represented and the functions to be performed" and have enough autonomy from the appointing power (Congress, the President, or an

28. See generally LUCIANO FLORIDI, SYLVIE KAUFFMAN, LIDIA KOLUCKA-ZUK, FRANK LA RUE, SABINE LEUTHEUSSER-SCHNARRENBERGER, JOSÉ-LUIS PIÑAR, PEGGY VALCKE & JIMMY WALES, REPORT TO THE ADVISORY COUNCIL TO GOOGLE ON THE RIGHT TO BE FORGOTTEN (2015).

29. Federal Advisory Committee Act, 5. U.S.C. § 2(a) (stating "they are frequently a useful and beneficial means of furnishing expert advice, ideas, and diverse opinions to the Federal Government" but imposing conditions).

agency head) to limit undue influence.³⁰ These constraints ensure FACA committees are independent and representative of various stakeholders. FACA also requires transparency in the information committees rely on and their decision-making processes. Most committee meetings must be noticed in the Federal Register and open to the public,³¹ and committee materials must be made available for public inspection.³²

Google's Advisory Council, in turn, consisted of eight members³³ selected by the company but with "no contractual relationship with Google on this project."³⁴ To further signal the council's independence from Google, Google did not require the GAC members to sign non-disclosure agreements and only reimbursed them for travel costs associated with the public and private meetings necessary for the project.³⁵ To support the work, the GAC reportedly relied on a range of documents, including non-confidential "publicly available" information, European Court of Human Rights case law, policy guidelines of news organizations, and the Article 29 Working Party's Implementation Guidelines. The GAC solicited expertise from Google and outside experts. Google provided the GAC with briefings from three experts: "an engineer, who explained Search; a Google lawyer,³⁶ who explained their compliance procedures; and a lawyer from an outside law firm, who explained the legal basis of the Ruling."³⁷ To gather additional expert opinions and stakeholder input, the GAC held seven public consultations with experts in Europe and gathered expert and lay opinions through a website.³⁸ In addition to these

30. 5 U.S.C. § 2(b).

31. 5 U.S.C. § 10(a)(2).

32. 5 U.S.C. § 11.

33. The eight members are Luciano Floridi, Professor of Philosophy and Ethics of Information at the University of Oxford; Sylvie Kauffman, Editorial Director, Le Monde; Lidia Kolucka-Zuk, Director of the Trust for Civil Society in Central and Eastern Europe; Frank La Rue, UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression; Sabine Leutheusser-Schnarrenberger, former Federal Minister of Justice in Germany; José-Luis Piñar, Professor of Law at Universidad CEU and former Director of the Spanish Data Protection Agency (AEPD); Peggy Valcke, Professor of Law at University of Leuven; Jimmy Wales, Founder and Chair Emeritus, Board of Trustees, Wikimedia Foundation. *Read the Advisory Council's final report*, GOOGLE ADVISORY COUNCIL, <https://archive.google.com/advisorycouncil>.

34. Jean-Marie Chenou & Roxana Radu, *The "Right to Be Forgotten": Negotiating Public and Private Ordering in the European Union*, 58 BUS. & SOC'Y 74, 90 (2019).

35. *Id.*

36. Daphne Keller was then Associate General Counsel for Intermediary Liability at Google and is now Director of Intermediary Liability at Stanford Law School's Center for Internet and Society.

37. FLORIDI ET AL., *supra* note 28, at 2. It is unclear whether the engineer was a Google employee or not. Google also provided staff support. *Id.*

38. *Id.* at 1-2.

public consultations, the GAC held three council-only meetings in which they deliberated and formulated guidance.³⁹ The GAC's final report, published on February 6, 2015, discussed alternative proposals that testifying experts offered at the seven public consultations; however, it did not reference the public comments received.⁴⁰ There is no public record of those comments.⁴¹

The GAC final report framed its guidance as addressing the right to “delisting”—the removal of “links returned in search results based on an individual’s name when those results are ‘*inadequate, irrelevant or no longer relevant, or excessive.*’”⁴² This was required under the *Google Spain* decision interpreting search engines obligations under Article 14 of the EU Data Protection Directive and the exceptions to this right where there is an overriding public interest in results “for particular reasons, such as the role played by the data subject in public life.”⁴³ Of particular importance to the report’s guidance is the conclusion that “whether the data subject experiences harm” from inclusion in a name-based search results page is “relevant to [the] balancing test” required to determine exceptions to the delisting right.⁴⁴ This conclusion appears to be at odds with the *Google Spain* ruling which makes no mention of a harm assessment and states, “it is not necessary in order to find such a right [to delisting] that the inclusion of the information in question in the list of results causes prejudice to the data subject.”⁴⁵ To inform its understanding of harm,⁴⁶ the GAC drew on case law addressing rights to data protection, privacy, and freedom of expression and information from the Court of Justice for the European Union (CJEU) interpreting the Charter of Fundamental Rights of the European Union⁴⁷ and the European Court of Human Rights (ECHR) interpreting the European Convention on Human Rights.⁴⁸ Based on the GAC’s analysis, the report concluded that “[t]he ruling, while reinforcing European citizens’ data protection rights, should not be interpreted as a legitimization for practices of

39. *Id.*

40. *Id.* at 34–37.

41. Chenou & Radu, *supra* note 34, at 88 (“[T]he comments submitted in response to the Request for Comments form on the Advisory Council’s website have not been published.”).

42. FLORIDI ET AL., *supra* note 28, at 2 (citing *Google Spain*, at ECLI:EU:C:2014:317, ¶ 94).

43. *Google Spain*, at ECLI:EU:C:2014:317, ¶ 97.

44. FLORIDI ET AL., *supra* note 28, at 6.

45. *Google Spain*, at ECLI:EU:C:2014:317, ¶ 96.

46. *Id.* at ¶ 97.

47. The Charter of Fundamental Rights of the European Union establishes the right to privacy (article 7) the right to data protection (article 8) and the right to freedom of expression and information (article 11). 55 O.J. (C 326) 391, 397–98.

48. The European Convention on Human Rights establishes the right to privacy (article 8) and freedom of expression (article 10). European Convention on Human Rights arts. 8, 10, Nov. 4, 1950 (amended 1998).

ensorship of past information and limiting the right to access information.”⁴⁹

The introduction of harm assessment restructures the test set out in the *Google Spain* ruling from one that looks at exceptions to data subjects’ rights necessitated by the “preponderant interest of the general public” (an inquiry looking at specific facts, such as the subject’s role in public life) to one that simply contrasts the harms to data subjects against the “preponderant interest of the general public.”⁵⁰ The GAC recommended that Google consider four key criteria when evaluating delisting requests: (1) the data subjects role in public life; (2) the nature of the information; (3) the source of the information, including its motivation for publication; and (4) time, “the notion that information may at one point be relevant, but as circumstances change, the relevance of that information may fade.”⁵¹ While the criteria and sub-criteria set out by the GAC capture many of the key concepts set out by the Article 19 Data Protection Working Party in their guidance, the introduction of balancing harms to data subjects against the public interest could substantially alter the outcomes achieved despite those shared criteria.

2. Facebook’s Oversight Board

The most recent symbolic legal structure to emerge is Facebook’s Oversight Board, a body explicitly likened to the most venerated of legal structures—the Supreme Court. In a trial balloon of the concept, Mark Zuckerberg said, “You can imagine some sort of structure, almost like a Supreme Court, that is made up of independent folks who don’t work for Facebook, who ultimately make the final judgment call on what should be acceptable speech in a community that reflects the social norms and values of

49. FLORIDI ET AL., *supra* note 28, at 6.

50. *Id.* at 5–6 (citing *Google Spain*, at ECLI:EU:C:2014:317, ¶ 97).

51. *Id.* at 7–14. The report went beyond the mandate to advise Google on decision-making criteria and included recommendations on the processes and inputs for decision-making. *Id.* at 15–21. It also included a discussion of the alternative ideas and technical proposals to establish adjudication processes presented during public consultations. *Id.* at 34–37. The alternative proposals discussed attend to democratic deficits inherent to the regulatory framework—as opposed to those that Google could address through implementation choices—including the lack of an administrative appeals process for users whose content was removed and the reliance on private rather than public processes to adjudicate claims. Among other proposals, experts suggested establishing “a clear channel of appeal to a public authority for publishers seeking vindication of Article 10 rights, parallel to data subjects’ right of appeal to DPAs,” and “a public mediation model, in which an independent arbitration body assesses removal requests,” modeled on the domain name dispute resolution process. *Id.* at 36.

people all around the world.”⁵² Zuckerberg explained that he had “come to believe that Facebook should not make so many important decisions about free expression and safety on [its] own.”⁵³

Facebook established the FBOB after receiving mounting criticism from policymakers, journalists, and the public regarding its decisions to remove and maintain content. As described above, many legal frameworks delegate substantial *definitional* work to platforms. In addition, many legal frameworks delegate responsibility for the *application* of the rules, and the processes and tools used to do so, to private platforms. The court analogy Zuckerberg chose responded to specific concerns with the *application* subfunction of the content moderation task.

In November 2018, Mark Zuckerberg announced a “blueprint” for increasing transparency and accuracy in content removals.⁵⁴ He outlined a number of internal changes, including improving Facebook’s efforts to independently identify and remove content in violation of Facebook’s community standards rather than relying on reports from users. However, the most significant change Zuckerberg announced was the creation of an independent oversight board to handle at least some content removal appeals.⁵⁵ To solicit guidance on the creation and function of the board and develop stakeholder buy-in, Facebook undertook an extensive consultation process involving public for a,⁵⁶ smaller expert working groups, as well as town halls.⁵⁷

52. Ezra Klein, *Mark Zuckerberg on Facebook’s hardest year, and what comes next*, VOX (Apr. 2, 2018, 6:00 AM EST), <https://www.vox.com/2018/4/2/17185052/mark-zuckerberg-facebook-interview-fake-news-bots-cambridge>.

53. Mark Zuckerberg, *A Blueprint for Content Governance and Enforcement*, FACEBOOK (Nov. 15, 2018), <https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634>.

54. *Id.* Peter Stern, Head of Product Policy Stakeholder Engagement, has indicated that the removal of the “After the ‘Terror of War’ ” and the ensuing controversy, was the impetus for these sweeping reforms. See Kate Klonick, *Facebook v. Sullivan*, KNIGHT FIRST AMEND. INST. (Oct. 1, 2018), <https://knightcolumbia.org/content/facebook-v-sullivan> (reporting on an interview with Peter Stern).

55. Kate Klonick exhaustively documented the path to the creation of the Board. See generally *The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression*, 129 YALE L. J. 2418 (2020).

56. See Brent Harris, *Getting Input on an Oversight Board*, META (Apr. 1, 2019), <https://about.fb.com/news/2019/04/input-on-an-oversight-board/> (describing public consultation process); Klonick, *supra* note 52, at 2448–57 (describing the consultation process as well as internal processes around the creation of the Board).

57. See Brent Harris, *Global Feedback and Input on the Facebook Oversight Board for Content Decisions*, META (June 27, 2019), <https://about.fb.com/news/2019/06/global-feedback-on-oversight-board/> (describing an entire process, which included expert consultations and public

The creation of the FBOB was Facebook’s effort to address the largely unchecked discretion and lack of transparency around processes used to enforce its own community standards. Facebook established the FBOB under the shadow of Section 230. Pursuant to the statute’s content moderation regime, platforms need not notify those whose content is subject to moderation; they need not respond to requests for moderation; they need not provide the reasons for moderation decisions; they need not provide any means to challenge content moderation activities; and they need not provide information about the processes, tools, and rules (if they exist) that guide their decision-making.

The wide latitude that platforms like Facebook have to moderate content under Section 230 permitted a lack of transparency about the rules and their application that were criticized as lacking the hallmarks of substantive and procedural legitimacy associated with adjudicating disputes in the public sector. Neither the parties nor the public were privy to the rules Facebook applied, the kinds of individuals or technology tasked with the application, or the controlling processes. Societal stakeholders raised concerns about Facebook’s lack of independence, how various incentives might influence their application of rules, and the lack of processes to contest content moderation decisions.⁵⁸

In establishing the FBOB, Facebook emulated the most symbolic of legal structures. The FBOB borrows features that are evocative of courts. The nine-page charter establishing the board contains detailed sections about membership, scope of authority, board procedures, implementation, board governance amendments and bylaws, and compliance with law.⁵⁹ Both Facebook and users can appeal to the FBOB for review of a content moderation decision, but the board retains discretion over which cases to hear. The charter establishes that the FBOB’s task is to consider whether decisions were “consistent with Facebook’s content policies and values” and that FBOB decisions set precedent. In the charter, Facebook committed to upholding the

consultation, and releasing report on the construction of the oversight board); Klonick, *supra* note 53, at 2448–57 (describing the consultation process as well as internal processes around the creation of the Board).

58. See, e.g., *The Santa Clara Principles On Transparency and Accountability in Content Moderation*, SANTA CLARA 1.0, <https://santaclaraprinciples.org> (last visited Mar. 18, 2022) (operational principle three calling on platforms to online speech platforms to create “meaningful opportunity for timely appeal” of moderation decisions and consider establishing “independent external review processes”).

59. *Oversight Board Charter*, FACEBOOK, https://about.fb.com/wp-content/uploads/2019/09/oversight_board_charter.pdf (last visited Mar. 18, 2022) [hereinafter *Oversight Board Charter*].

rulings of the Board as final unless it would violate the law to do so.⁶⁰ The charter also contains several provisions to bolster the FBOB's independence.⁶¹ In May of 2020, Facebook announced the first members of the Board.⁶² Board members included individuals with "experience in press freedom, digital rights, religious freedom, content moderation, online safety, internet censorship, platform transparency and technology."⁶³

Together, the lack of transparency into how platforms apply rules to specific content, the absence of notice and participation rights for relevant parties, and the lack of a public record of the reasoning behind determinations have undermined the perceived legitimacy of platforms' application of content moderation standards. The court-like aspects of the FBOB evince a direct attempt to gain legitimacy by adopting structures and processes that support the independence of decision makers, transparency of decisions in the particular case, and affected parties' participation in public adjudicatory processes.

3. *Transparency Reports*

Over the last twelve years, social media platforms and other information and communication technology companies have begun releasing "transparency reports," which share information about content removals and disclosures of users' personal information.⁶⁴ Google released the first transparency report in 2010, and many other technology companies have adopted some version of transparency reporting in the intervening years.⁶⁵

60. Catalina Botero-Marino, Jamal Greene, Michael W. McConnell & Helle Thorning-Schmidt, *We Are a New Board Overseeing Facebook. Here's What We'll Decide*, N.Y. TIMES (May 6, 2020), <https://www.nytimes.com/2020/05/06/opinion/facebook-oversight-board.html>.

61. The Board is funded through an independent trust, which was set up by Facebook but cannot be revoked by the company. Each board member will serve fixed terms of three years and may serve up to three terms. *Announcing the First Members of the Oversight Board*, OVERSIGHT BD. (May 6, 2020), <https://www.oversightboard.com/news/announcing-the-first-members-of-the-oversight-board/>.

62. Nick Clegg, *Welcoming the Oversight Board*, META (May 6, 2020), <https://about.fb.com/news/2020/05/welcoming-the-oversight-board/>. In August of 2019, Facebook had issued further guidance, explaining how the Board would operate. *Facebook Oversight Board for Content Decisions: What to Know*, META (Aug. 22, 2019), <https://www.facebook.com/journalismproject/facebook-oversight-board-for-content-decisions-overview>.

63. *Id.*

64. Peter Micek & Isedua Oribhabor, *The what, why, and who of transparency reporting*, ACCESS NOW (Apr. 2, 2020), <https://www.accessnow.org/the-what-why-and-who-of-transparency-reporting/>.

65. See James Losey, *Surveillance of Communications: A Legitimization Crisis and the Need For Transparency*, 9 INT'L J. COMM'N 3450, 3453 tbl.1 (2015) (providing an overview of Transparency Reports of forty-one information and communication technology sector

Companies initially adopted transparency reports to increase public visibility into government requests for the personal information and communications of platform users⁶⁶ and, for some platforms, content removals made at the behest of third parties.⁶⁷ The reports adopted the form of wiretap reports, which the U.S. Department of Justice files yearly detailing the number and categories of state and federal wiretaps requested and issued.⁶⁸ By mirroring a practice used in public governance, companies sought to “reassure national and international subscribers that the[y] had rigorous processes for evaluating government requests for data, shed light on the

companies, documenting that some publish data about removals in multiple countries, all provide data about demands for personal data, but not all publish data about content removal, and to the extent they do the data varies in coverage); *Transparency Reporting Index*, ACCESS NOW, <https://www.accessnow.org/transparency-reporting-index/> (last visited Apr. 20, 2022). The earliest effort to provide transparency about content removed by social media platforms was the Chillingeffects.org website. Maintained by a set of law school clinics, the website was a repository for DMCA takedown notices. In 2003, Google began contributing the takedown notices and, importantly, providing a link to the site where relevant results had been removed. A few smaller internet service providers contributed takedown notices as well. However, more recently some companies have “quietly dropped the practice,” and “not a single household-name tech firm seems to have adopted [the reports] since early 2016.” Rob Pegoraro, *Tech Companies Are Quietly Phasing Out a Major Privacy Safeguard*, THE ATLANTIC (Sept. 29, 2019), <https://www.theatlantic.com/technology/archive/2019/09/what-happened-transparency-reports/599035/>.

66. The first Ranking Digital Rights Corporate Accountability Index published in 2015 shows that the sixteen companies evaluated were generally providing more transparency into government requests for customer data than government requests to remove content. *Compare P11. Data about third-party requests for user information*, RANKING DIGIT. RTS., <https://rankingdigitalrights.org/index2015/indicators/p11/> (last visited Apr. 20, 2022), *with F7. Data about Government Requests*, RANKING DIGIT. RTS., <https://rankingdigitalrights.org/index2015/indicators/f7/> (last visited Apr. 20, 2022).

67. Google provides a brief history of the development of their transparency reports at GOOGLE TRANSPARENCY REP., <https://transparencyreport.google.com/about?hl=en> (last visited Apr. 20, 2022). From inception, they were designed to provide information about both demands for user data and government requests for content removal. In 2009, Google had been blocked in 25 different countries but wanted to reveal that short of full blocking governments were demanding the removal of specific content. Nicole Wong, *Dinner Speech at Conference on Liberation Technology in Authoritarian Regimes 4* (Oct. 11, 2010), https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/evnts/media/2010-10_Nicole_Wong_Stanford_Liberation_Technology.pdf (“[W]e wanted to bring some transparency to what is certainly only our limited view on government activity on the Internet. It is not complete. It is not complete because it is only about our products. It is not complete because our data is not sufficiently granular enough yet. It is not complete because some governments will not even let us publish this data. So, why did we do it? The conversation about government censorship and surveillance has to start somewhere.”).

68. *See, e.g., Wiretap Report*, U.S. CTS., <https://www.uscourts.gov/statistics-reports/wiretap-report-2019> (Dec. 31, 2019) (providing information on wiretaps by jurisdiction and crime, as well as other information).

regularity and breadth of such requests, and encourage reforms in government surveillance activities.”⁶⁹

Since 2010, the use and scope of transparency reports have expanded in two ways. First, they now provide information about content removals in response to legal processes by private parties as well as government entities,⁷⁰ shedding light on what legal scholar Jack Balkin has called “new-school speech regulation,” by which owners of platforms and other digital infrastructure are “coerce[d] or co-opt[ed]” into regulating speech.⁷¹

Second, in response to public criticism of content removals under platforms’ Terms of Services,⁷² the Reports increasingly include information about

69. Christopher Parsons, *The (in) Effectiveness of Voluntarily Produced Transparency Reports*, 58 BUS. & SOC’Y 103, 112 (2019); see, e.g., *Twitter Transparency Center*, TWITTER, <https://transparency.twitter.com/en/about.html> (last visited Mar. 18, 2022) (“The original goal of our transparency report was to provide the public with recurring insights into government pressures that impacted the public.”).

70. Beginning in 2012 Google began providing data about content removals pursuant to DMCA requests. As Daphne Keller and Paddy Leersen discuss, current transparency reports do not provide data about removal requests under all laws. DAPHNE KELLER & PADDY LEERSEN, *FACTS AND WHERE TO FIND THEM: EMPIRICAL RESEARCH ON INTERNET PLATFORMS AND CONTENT MODERATION*, in *SOCIAL MEDIA AND DEMOCRACY: THE STATE OF THE FIELD, PROSPECTS FOR REFORM* 220, 228 (N. Persily, & J. A. Tucker eds., 2020) (“[M]ost transparency reports only cover particular categories of takedowns—often only those initiated by governments or copyright-holders. This leaves open questions about platforms’ responses to legal allegations brought by individuals under, say, French defamation law or Brazilian privacy law.”).

71. Jack M. Balkin, *Free Speech is a Triangle*, 118 COLUM. L. REV. 2011, 2016–17 (2018) (describing risks of new-school speech regulation: collateral censorship and digital prior restraint). This effort was also informed by an earlier effort to highlight the behavior of copyright holders wielding the power of the DMCA: the Chilling Effects database founded in 2002. LUMEN, <https://lumendatabase.org/> (last visited Mar. 18, 2022). That database, a collaborative project established by a group of law school clinics and the Electronic Frontier Foundation, collected copies of takedown requests that were contributed by recipients, Google, and some smaller service providers; annotated and archived them; and made them available for retrieval through search engines. Google, an initial contributor of notices, included a notice about search results removed due to DMCA requests at the bottom of their search results page and provided a link to the relevant takedown request in the Chilling Effects database. This infrastructure provided web searchers and researchers insight into information removed in response to DMCA requests. Chilling Effects was replaced by Lumen and is run by the Berkman-Klein Center for Internet and Society. *About Us*, LUMEN, <https://lumendatabase.org/pages/about> (last visited Apr. 20, 2022).

72. These include Facebook’s removal of a Pulitzer Prize winning graphic photo of a naked Vietnamese girl suffering as napalm from a U.S. attack burned her skin (The Terror of War), ongoing concerns about biases of all sorts in corporate content moderation, and the government use of terms of service violations as a quick and invisible way to remove content, including content they may be unable to remove through legal processes. Zoe Kleinman, *Fury over Facebook ‘Napalm girl’ censorship*, BBC NEWS (Sept. 9, 2016), <https://www.bbc.com/news/>

content removals under their terms of service. In 2015, Ranking Digital Rights, which annually evaluates company policies and practices affecting speech and privacy, reported that no company was regularly publishing data about content moderation or account suspensions based on the company's terms of service or other rules.⁷³ By 2019, for example, Facebook, Google, and Twitter disclosed "comprehensive data about content removals due to terms of service enforcement" and Microsoft was publishing some information, although less comprehensive, about terms of service enforcement.⁷⁴

Today, social media platforms use transparency reports to provide data on the extent and kind of government queries for user data (privacy); the extent, kind, and action platforms take in response to requests for content removal by government and private parties; and content removals under their terms of service. Twitter's Transparency Center—an interactive online version of its original transparency reports—reflects this evolution. The Center now provides relatively robust data and visualizations about the company's content moderation activities.⁷⁵ Twitter includes data about external requests for data and internal content moderation activities taken in response to a wide range of laws as well as Twitter's rules. The Center provides jurisdiction-specific data, an interactive tool for comparing countries against each other, data about trends of across time and jurisdictions, and examples offering detail and

technology-37318031; Espen Egil Hansen & Dear Mark, *I Am Writing This to Inform You that I Shall Not Comply with Your Requirement to Remove This Picture*, AFTENPOSTEN (Sept. 8, 2016, 9:33 PM), <https://www.aftenposten.no/meninger/kommentar/i/G892Q/dear-mark-i-am-writing-this-to-inform-you-that-i-shall-not-comply-wit>; Michael Nunez, *Former Facebook Workers: We Routinely Suppressed Conservative News*, GIZMODO (May, 9, 2016), <https://gizmodo.com/former-facebook-workers-we-routinely-suppressed-conser-1775461006>; see Brian Chang, *From Internet Referral Units to International Agreements; Censorship of the Internet by the UK and EU*, 49 COLUM. HUM. RTS. L. REV. 114 (2017).

73. RANKING DIGIT. RTS., 2019 RDR CORPORATE ACCOUNTABILITY INDEX 42 (2019) [hereinafter RDR 2019], <https://rankingdigitalrights.org/index2019/assets/static/download/RDRindex2019report.pdf>.

74. *Id.* Twitter was the first to disclose information about actions taken under its Terms of Service in its transparency report. RDR 2019 also found that all companies reviewed disclosed basic information about their terms of service. *Id.* at 44. The most recent version of Ranking Digital Rights' Corporate Accountability Index, reviewing twenty-six companies, can be found at *The 2020 RDR Index*, 2020 RANKING DIGIT. RTS. CORP. ACCOUNTABILITY INDEX, <https://rankingdigitalrights.org/index2020/> (last visited Apr. 20, 2022). Google maintains a non-exhaustive list of entities publishing transparency reports at GOOGLE TRANSPARENCY REP., <https://transparencyreport.google.com/> (last visited Apr. 20, 2022).

75. *Twitter, Inc.*, 2020 RANKING DIGIT. RIGHTS CORP. ACCOUNTABILITY INDEX, <https://rankingdigitalrights.org/index2020/companies/Twitter> (last visited Apr. 19, 2022) (finding that Twitter shared more data about the enforcement of platform rules than its peers and that it discloses more data about government demands for content removal and user information than most of its U.S. peers; ranking Twitter #1, with a score of 53% on the Index).

context about specific actions.⁷⁶ Yet, despite the steady expansion, transparency reports do not provide a complete picture of content removals or content moderation practices⁷⁷ or requests for personal information.⁷⁸

B. SHIFTING THE FOCUS FROM FORM TO FUNCTION

By engaging the public, bringing in outside experts, and producing records of decision-making in forms that resemble those of traditional legal institutions, platforms have attempted to legitimize substantive outcomes through the adoption of processes associated with the legitimate institutional exercise of legal power. Yet commentators have expressed a deep sense that “[t]he inwards-looking, largely public relations-oriented content governance models so widely deployed today are unsatisfying.”⁷⁹

76. See, e.g., *Removal Requests*, TWITTER TRANSPARENCY (Dec. 2020), <https://transparency.twitter.com/en/reports/removal-requests.html#2020-jul-dec> (finding that through the “removal requests” report section a visitor to the Transparency Center can view worldwide statistics and a short analysis section noting that Japan accounts for 43% of global requests and that those requests are primarily related to laws regulating control substances, obscenity, and money lending.) The visitor could then select the Japan specific report and review more data. *Japan*, TWITTER TRANSPARENCY, <https://transparency.twitter.com/en/reports/countries/jp.html> (last visited Mar. 18, 2022). Finally, a visitor could compare Japan’s statistics to any one of the other one hundred and nine countries for which the Center provide reports or with [external] worldwide data. *Id.*

77. *F4a. Data about content restrictions to enforce terms of service*, 2020 RANKING DIGIT. RTS. CORP. ACCOUNTABILITY INDEX, <https://rankingdigitalrights.org/index2020/indicators/F4a> (last visited Apr. 19, 2022) (reporting that Facebook and Google do not publish data about the total number of pieces of content restricted for violating the company’s rules, Twitter provides partial data on this topic, and describing other limitations of current reports); Hannah Bloch-Wehba, *Global Platform Governance: Private Power in the Shadow of the State*, 72 S. METHODIST U. L. REV. 27, 72–74 (2019) (describing limitations in transparency reports including, lack of information about the number of videos Google removed for “violent extremism,” lack of clarity about what actions Facebook has taken with respect to various pieces of content, and lack of standardization).

78. See Hannah Bloch-Wehba, *Exposing Secret Searches: A First Amendment Right of Access to Electronic Surveillance Orders*, 93 WASH. L. REV. 145, 158–62 (2018) (providing an overview of how the law constrains what the public knows and what platforms can report about government requests for user data); Alex Abdo, *More Transparency Needed For Government’s Use of National Security Powers For Data Requests From Companies*, AM. C.L. UNION (June 19, 2012) (explaining that Google’s—and other companies—2011 transparency report provided no insight into the use of surveillance authorities, such as National Security Letters, to obtain user data). Companies can now provide information about national security letters and FISA orders in bans of 250. Letter from James M. Cole, U.S. Deputy Attorney Gen., to General Counsels of Facebook, Google, LinkedIn, Microsoft, and Yahoo, January 27, 2014 (on file with authors).

79. John Bowers & Jonathan Zittrain, *Answering Impossible Questions: Content Governance in an Age of Disinformation*, 1 HARV. KENNEDY SCH. MISINFO. REV. 1, 5 (2020).

Assessing whether these symbolic structures enhance the legitimacy of the content moderation function—by directly addressing democratic deficits and promoting public values—poses a real challenge. How can we tell whether they are merely symbolic⁸⁰—providing companies with legitimacy without addressing underlying concerns—or examples of a salutary “[p]rocess era of internet governance”⁸¹ that reflects legitimate models for addressing controversial content governance issues that pit individual rights against collective interests?

Simply weighing the negative and positive attributes of individual structures offers little purchase on the question; these questions confound an easy answer at such a high level of generality. The task of content moderation is not an undifferentiated soup in which a dash of “transparency,” “participation,” or “fairness” can provide the right taste. Assessing whether or these structures “build legitimacy around how content is sorted, filtered, and ranked”⁸² requires greater clarity about the standard(s) against which to measure discrete interventions.⁸³

A closer look at the foregoing descriptions of the three legality-evoking forms points to a different lens for analyzing the different structures employed by platforms engaged in content moderation. Specifically, rather than assessing these structures as tools that seek to legitimate content moderation function *as a whole*, they in fact represent attempts to address deficits in particular tasks, or subfunctions, that together comprise the broader function.

Understood in this way, Google’s Advisory Council was not assigned the power to conduct all aspects of content moderation but to add legitimacy to one subfunction: *defining* rules and policies governing content. The legal framework created liability but gave next to no guidance on how to evaluate the privacy claims of individuals, the competing interests of the public and other stakeholders, or any guidance on how to weigh the two. Absent such guidance, Google was implicitly tasked with the complex and contentious definitional work of moderating content. They attempted to legitimate the definitional and operational work necessary to moderate content by assigning

80. Edelman, *supra* note 8, at 1542.

81. Bowers & Zittrain, *supra* note 77, at 7.

82. *Id.* at 5.

83. Barrie Sander parses out different content moderation activities along the lines of their purpose to support human rights analysis and due diligence. *See* Sander, *supra* note 10, at 998–90 (identifying four platform content moderation activities in which transparency and oversight should be provided: rule making, decision-making, content and advertising, and regulatory compliance).

that subfunction to a diverse group of independent experts deliberating in a transparent and public process.

Facebook assigned to its Oversight Board, in turn, authority over a second content moderation subfunction: the *application* of rules and policies to determine whether specific content meets the platform's governing definition. Such a task involves both construing content in context to determine its meaning and interpreting platform rules and decisional criteria to decide their application. Thus, through the FBOB, Facebook has attempted to demonstrate consistency with some of the core competencies involved in traditional forms of law-application through public adjudication.

Transparency reports, in their current form, provide public transparency into the outcomes of a third subfunction: case *resolution*. Unlike actions taken through court processes, there is limited visibility into the information removed from the web through the processes emerging under today's content moderation regimes. Transparency reports seek to empower a variety of stakeholders by creating a public record of decisions to remove expression from the public view.

Looking at content moderation through this functional lens reflects the reality of the whole range of statutory frameworks that deal with the treatment of content online. While Section 230 provides no explicit guidance regarding the allocation of subfunctions to different actors, numerous other laws reflect an intention by Congress to do so explicitly, in a variety of ways.

Understanding the symbolic forms adopted by platforms through a functional lens suggests new questions to ask when assessing their use in content moderation:

To what extent (if at all) does Google's enlistment of input from the public into the *definition* of rules and policies governing content provide competencies that can address democratic deficiencies in that subfunction?

To what extent (if at all) does Facebook's assignment of oversight over the *application* of rules in specific situations to experts address concerns about values raised by that subfunction?

To what extent (if at all) does the information about the *resolution* of specific cases provided through transparency reports enlist competencies that remedy deficits raised by that subfunction?

With these questions in mind, the next Part isolates public values attached to the three subfunctions already identified—definition, application, resolution—and adds a fourth: *identification*. It then examines ways that different tasks or subfunctions are allocated to particular actors (private and

public; human and technical) and the competencies that each contribute to achieving those values through the multi-actor system as a whole. Part III will then return to the assessment of the symbolic structures discussed above, using the analytic framework presented below to assess these legalistic forms, and suggest ways that legislative frameworks might better structure corporate content moderation.

III. DEVELOPING A FUNCTIONAL FRAMEWORK

Informed by the understanding that content moderation does not involve a single function, but rather a set of discrete subfunctions, this Part describes four discrete subfunctions. The subfunctions identified are common across diverse content moderation systems, although the choices made about their assignments and structures vary. Specifically, this functional framework identifies the set of discrete subfunctions as:

- (1) the *definition* of the content subject to moderation;
- (2) the *identification* of potentially covered content;
- (3) the *application* of the definition to identified content; and
- (4) the *resolution* of a particular case (including labeling, amplifying, depressing, or removing).

Distinguishing these subfunctions at a granular level permits a more rigorous inquiry into which actor, or combination of actors, might best be tasked with their performance and how the performance of those subfunctions should be structured. This helps pinpoint which governance values are most salient to different stages of the content moderation process; what competencies are required to perform the subfunction consistent with those values; which actor or combination of actors might provide those competencies; and what constraints should be imposed to ensure that those competencies are brought to bear.

This Part uses existing regulatory frameworks to illustrate common subfunctions, variations in their implementation, and ways policy choices shape them. The examples, drawn from Section 230, the Digital Millennium Copyright Act, the General Data Protection Regulation, and the mix of regulations that shape how platforms handle child sexual abuse material, together illuminate the choices policymakers and other stakeholders can make to guide and constrain decision-making using each subfunction. This analysis seeks to identify regulatory frameworks that allocate and constrain content moderation subfunctions to be more or less supportive of the democratic values bound up in the regulation of speech.

A. GROUNDING A FUNCTIONAL APPROACH

Our suggestion that a rigorous assessment of content moderation must focus on the component subfunctions, the different ways those subfunctions are allocated and constrained, and the deficits that arise under them that reflect an absence of governance-related competencies, is grounded in theoretical insights from “New Governance,” values-in-design scholarship, and the realities of the current content moderation practices of platforms.

1. *Theoretical Roots*

A functional approach reflects the insights of our and others’ work on “New Governance.”⁸⁴ This approach recognizes the ways that legal frameworks—from detailed imposition of content standards to liability regimes—delegate, explicitly or implicitly, the content moderation function to private platforms.⁸⁵ Accordingly, it focuses on looking both *within* and *across* the black box of networked-organizational decision-making processes responsible for content moderation on a granular level. The goal is to further the alignment of these processes with public governance norms,⁸⁶ by taking seriously the choices between distinct human and technical actors within those contexts and the networks they engage.⁸⁷ It further draws attention to emerging trans-governmental networks and transnational forms of private regulation, many of which delegate functions of content moderation to technical actors.⁸⁸ Extending the focus to private governance activities, and their comportment with public values, can in turn “catalyze the ongoing development of meaningful internal practices.”⁸⁹

More specifically, this functional orientation to regulatory allocations in content moderation applies the “handoff model” that one of the authors has

84. See, e.g., Bamberger & Mulligan, *supra* note 20, 480–82 (summarizing the “new governance” literature).

85. See Bamberger, *Regulation as Delegation*, *supra* note 20, at 383 (describing how “private firms increasingly exercise regulatory discretion of the type delegated to agencies”).

86. See *id.* at 384 (suggesting replacing a “compliance” paradigm for one focused on “accountability”); *id.* at 383 (proposing the need for a richer account of decision-making within the corporate “black box” to understand the extent to which firms’ exercise of regulatory discretion is accountable to public norms).

87. See, e.g., Bamberger, *Technologies of Compliance*, *supra* note 20, at 673 (focusing on the use of technology systems in risk management decision-making).

88. See, e.g., Gregory Shaffer, *Theorizing Transnational Legal Ordering*, 12 ANN. REV. L. & SOC. SCI. 231, 239 (2016) (discussing the ways that the delegation of content moderation functions to technical actors, at times placing “public law . . . in the shadow of transnational private regulation”).

89. Bamberger & Mulligan, *supra* note 20, at 482.

developed with philosopher Helen Nissenbaum.⁹⁰ The handoff model suggests that the values implications of decision-making models cannot be understood by looking generally at an overall “function,” like content moderation itself. Rather, it requires isolation, separation, and examination of the subfunctions that comprise the content moderation function and of the ways that such subfunctions are assigned to differently situated humans, technologies, and combinations of the two.

Current debates in content moderation speak to the significance of the ways that platforms enact content moderation through these functional handoffs. For example, particular corporate implementations of the content moderation function have been critiqued for misidentifying content as covered by platform policy due to an inability to account for context that is essential to appropriately applying a definition.⁹¹ Both humans and technical actors have misidentified content, yet the cause of misidentifications stem from different limitations and biases. While scholars have often raised concerns about privatization and automation, the handoff model promotes a more precise exploration of *how* each actor enlisted in a content moderation regime executes the function delegated to it.

Dissecting subfunctions enacted under various content moderation regimes reveals important questions about the quality of performance, efficiency, and the effect on markets. More importantly, here, it directs attention towards the ethical and political questions that arise in alternate versions of content moderation systems. It complicates the external appearances of *sameness*, exposes how different allocations of content moderation functions implicate different values, and permits an analysis of the range of constraints that might accompany specific functions to protect important public values.⁹²

90. *The Concept of Handoff*, *supra* note 17; see also Kenneth W. Abbott & Duncan Snidal, *The Governance Triangle: Regulatory Standards Institutions and the Shadow of the State*, in *THE POLITICS OF GLOBAL REGULATION* 46 (Walter Mattli & Ngaire Woods, eds., 2009).

91. See, e.g., Aarti Shahani, *With ‘Napalm Girl,’ Facebook Humans (Not Algorithms) Struggle To Be Editor*, NPR: ALL TECH CONSIDERED, (Sept. 10, 2016), <https://www.npr.org/sections/alltechconsidered/2016/09/10/493454256/with-napalm-girl-facebook-humans-not-algorithms-struggle-to-be-editor> (discussing human reviewers removing historically significant photos that contained nudity under Facebook’s “community standards”); Louise Matsakis, *Tumblrs Porn-Detecting AI Has One Job—And It’s Bad At It*, WIRED.COM (Dec. 5, 2018), <https://www.wired.com/story/tumblr-porn-ai-adult-content/> (algorithms identifying and removing images as “adult content” under Tumblrs policy).

92. As an example, consider the application of a handoff lens to the access control technologies used in various generations of the Apple iPhone. See *The Concept of Handoff*, *supra* note 17, at 242–248. These have progressed from user-selected passwords to Touch ID (the fingerprint recognition system), to Face ID, by which the iPhone camera constructs a 3D map

The insights of regulatory scholars Kenneth Abbot and Duncan Snidal enhance this analysis. They suggest that evaluation of the strengths and weaknesses of multi-actor governance schemes should focus on whether the actors tasked with performing specific tasks possess, or can harness, necessary competencies.⁹³ Those competencies include independence, representativeness, expertise, and operational capacity, and different sets of them are essential to the legitimacy of distinct governance activities.⁹⁴ “It is difficult if not impossible,” they write, “for any non-state actor to provide all the competencies on its own. Thus, the most promising strategy may be collaboration: assembling the needed competencies by bringing together actors of different types.”⁹⁵ Assessing the capacities of the different actors to whom various subfunctions of content moderation are allocated offers a means to assess whether the complete suite of competencies has been brought to bear in a way that harnesses organizations’ and specific human and technical actors’ competencies and addresses relevant deficits in the emergent, networked, regulatory system.⁹⁶

2. *Descriptive Realities*

The functional framework further reflects the reality of existing content moderation practices. Expanding the category of “content moderation” to include the vast and diverse array of laws that involve the regulation of online

of a person’s face. A typical narrative for framing the substitution of these different mechanisms might emphasize technological progress: the *same* access function is being performed by increasingly sophisticated technological means, resulting in an upward linear trajectory in terms of security and, perhaps, user experience. The handoff lens’ emphasis on the systemic relation of both technological and human inputs into the system, by contrast, reveals that the choice of mechanism implicates important differences in terms of values, including human control and agency, transparency, and privacy.

93. Abbott & Duncan, *supra* note 90, at 44–88.

94. *Id.* at 46.

95. *Id.*

96. See Robert Gorwa, *The Platform Governance Triangle: Conceptualising The Informal Regulation Of Online Content*, 8 INTERNET POL’Y REV. 1, 13 (2019) (using Abbott and Snidal’s “governance triangle” to analyze a range of content governance schemes and noting the need for research focused on the varying regulatory competencies different actors bring to the content moderation).

material—including data protection,⁹⁷ civil rights,⁹⁸ intellectual property,⁹⁹ and other laws, as well as voluntary agreements that drive platforms to moderate content—reveals a host of content moderation practices that operate very differently from the canonical Section 230 framework. While Section 230 provides essentially no guidance regarding the moderation of content, these other regimes feature statutory schemes that are more explicit in allocating different subfunctions to different actors subject to a range of constraints. Focusing on content moderation as a functional output provides a set of in-the-wild case studies regarding the possibilities of more thoughtful assignment of different tasks to achieve different goals, leverage distinct competencies, and mitigate or backfill deficits.

B. THE FUNCTIONAL FRAMEWORK

The handoff model argues that dissecting content moderation into its component parts facilitates an understanding of the concerns about values raised at each step and the ways that different allocations of subfunctions to actors with different constraints can support or undermine values.

97. GDPR, 2016 O.J. (L 119) 1, http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf.

98. *See, e.g.*, Title VII of the Civil Rights Act of 1964 §§ 703–16, 42 U.S.C. § 2000e to 2000e-15 (prohibiting discrimination in employment); *id.* at § 2000e-3(b) (prohibiting advertisements that “indicate a preference, limitation, specification or discrimination” with respect to protected classes); Age Discrimination in Employment Act (ADEA), 29 U.S.C. § 623(e) (similar provision prohibiting advertisements that express a preference); Fair Housing Act (FHA), 42 U.S.C. § 3601 (prohibiting housing discrimination against protected classes); *id.* at § 3608 (prohibiting advertisements that “indicate a preference, limitation, specification or discrimination” with respect to protected classes).

99. Transparency reports produced by key platforms, as discussed in Section III.A.3, *infra*, include information about content moderation occurring under a range of intellectual property legal frameworks across jurisdictions, including copyright, trademark, and trade secret. *See, e.g.*, *Copyright Notices*, TWITTER TRANSPARENCY, <https://transparency.twitter.com/en/reports/copyright-notices.html#2020-jul-dec> (last visited Mar. 18, 2022) (providing separate reports on removals under the DMCA); *Trademark Notices*, TWITTER TRANSPARENCY, <https://transparency.twitter.com/en/reports/trademark-notices.html#2020-jul-dec> (last visited Mar. 18, 2022) (same under Twitter’s trademark policy). Google provides a specific transparency report on requests to delist links from search results based on copyright, and their resolution and their general content removal statistics include content removed “due to claims of trade dress and/or distinctive marks.” *Content Delistings due to copyright*, GOOGLE TRANSPARENCY REP., <https://transparencyreport.google.com/copyright/overview> (last visited Mar. 18, 2022). This includes, but is not limited to, claims of counterfeit and trademark. GOOGLE TRANSPARENCY REP. HELP CTR., <https://support.google.com/transparencyreport/answer/7347744?hl=en> (last visited Apr. 20, 2022). Facebook’s Transparency Center includes a section on actions taken on claims of intellectual property violations, including copyright, trademark, and counterfeit goods. *Intellectual Property*, META, <https://transparency.fb.com/data/intellectual-property/> (last visited Mar. 18, 2022).

A given legal scheme may not explicitly designate who is responsible for a particular subfunction. Yet in practice, a different actor or set of actors—public or private, human or technical—performs each subfunction and the allocations that emerge under existing content moderation regimes differ. For example, a social media platform could employ more reactive or proactive measures to initiate cases by relying on users to report potential violations of content policies, training employees to perform the identification task, using an automated system that relies on natural language processing, or a combination of all three.

At times, subfunctions may be encumbered with procedural constraints, such as a requirement to provide notice to a user when removing content, or constraints on the actors who can actually remove content, such as encouraging the use of an automated filter to screen out objectionable content or constraining the use of automation.

Below we define each discrete subfunction, note the core values associated with its performance and the competencies those demand, and use case studies to explore and contrast how different content moderation regimes allocate the subfunctions.

1. *Components of A Functional Framework*

a) Identifying the Subfunctions of Content Moderation

Every content moderation regime explicitly or implicitly assigns responsibility for a set of discrete subfunctions that comprise the content moderation task: definition, identification, application, and resolution.

Definition—Each content moderation regime must set rules or policies defining the type of content it targets. Different regimes afford distinct actors more influence, at least for some period, over the definition of content subject to moderation. The actor who is authorized to craft the definition wields immense power.

While it may be tempting to assume that formal law does the heavy lifting on this important aspect of content moderation policymaking, that is often not the case. A definition may be explicitly captured in statutory language or case law. Yet, even when a statutory definition exists, it may take the form of a multi-factor balancing test or a broad standard that in practice shifts power for defining content to the entity facing liability. Sometimes the decision about which content to moderate is left entirely to platforms, without any requirement that they formally create definitions, leaving the definition of content subject to moderation intuited through removals rather than through *ex ante* definitions or guidelines.

Identification—Once the content to be moderated is defined, some actor(s) must be tasked with initiating inquiries by identifying potentially covered content—the functional equivalent of bringing a legal case. This subfunction might be assigned to the platform itself (through direct commands or liability regimes), to public actors (such as prosecutors, law enforcement agents, and regulatory agencies), or to a range of other private actors (including rights-holders or other parties claiming injury. This range of responsible actors, further, may rely on human efforts to flag relevant content or technical systems using artificial intelligence.

The identification subfunction can also include the subsidiary task of actually locating the content online. In some instances, the process for identifying content includes identifying its location; reporting systems, for example, sometimes require provision of a URL.¹⁰⁰ Yet a content moderation regime might explicitly or implicitly assign the responsibility for identifying and locating content to distinct actors.¹⁰¹ For example, under the context of the National Center for Missing and Exploited Children database discussed below,¹⁰² once content is identified child sexual abuse material by human analysts, technical system—hash databases of that content—are used to support the location and summary removal of matching images. The ability to bifurcate identifying and locating creates interesting opportunities to leverage the competencies of distinct public and private, and human and technical, actors. Many contentious battles over online content center on whether platforms have a responsibility for identifying and locating regulable content.

Application—Once the content to be moderated is defined and potential instances of covered content is identified, a determination must be made as to whether the content meets the definition. This subfunction involves the application of a rule to a particular fact pattern, a function traditionally assigned to public processes of adjudication, whether judicial or administrative. Such application often involves both interpreting the connotations of the content in context, the meaning of the decisional rules, and the fit between them. Thus, depending on the type of content and relevant rules involved, the legitimacy of such a process requires not just a technocratic application of rules to facts but also constraints that ensure that judgment and interpretive discretion satisfies a range of public values, from democratic oversight, to participation, to consistency, proportionality, and fairness. Achieving this balance has posed

100. See, e.g., *infra* Section II.B.2.b (discussing the right-to-be-forgotten context).

101. In addition, the location of the content may alter the identification. For example, using a video of police brutality that has been identified in a hate group context to locate that video in a news context may alter the outcome of the identification subfunction.

102. See *infra* text accompanying notes 134–139.

an especially thorny task when authority is vested in supra-national, or private, decision-making bodies rather than regulatory or enforcement agencies or public prosecutors.¹⁰³

Different content moderation regimes assign this role to different actors and sometimes to multiple actors over the course of a dispute. Sometimes, as in the case of the Digital Millennium Copyright Act, public courts remain a key actor; in other regimes, such as Section 230, the private sector performs the application task entirely.

Resolution—This subfunction includes decisions about the full range of actions that may be taken once content is identified and located and the decisional rule applied. Resolving cases involves a determination of appropriate remedies. Sometimes the law explicitly or implicitly determines the action, by imposing strict liability for certain content or setting forth detailed statutory provisions directing removal. Other regimes leave platforms free to determine resolution decisions. Common moderating actions include blocking, removal, amplifying, downgrading, flagging, labeling, monetizing, strategically engaging, and reporting (including but not limited to government agencies), and the list continues to expand.¹⁰⁴ While the platform is the actor that most commonly comes to mind as having the right, obligation, or ability to resolve content issues, many platforms provide individual users with the ability to engage in at least some aspects of the task directly, for example enabling users to mute or block content from specific users, to mute specific messages, or to control whether they see content that a platform has designated offensive. Users, moreover, can also construct methods of resolving content issues, creating blocklists and tools that automate them through tools such as Block Bot and the now-defunct Block Together.

The order in which these subfunctions occur can differ by content moderation regime. Frequently, for example, disputed or problematic content is identified, and the rules and policies for determining whether they meet the relevant definition are then applied. But, as in the case of the CSAM moderation regime discussed below,¹⁰⁵ the process of determining that the definition applies to a particular image is sometimes made first, and instances of that image subsequently identified and resolution achieved, without

103. See, e.g., Martin Shapiro, “Deliberative,” “Independent” Technocracy v. Democratic Politics: Will the Globe Echo the E.U.?, 68 L. & CONTEMPORARY PROBLEMS 341 (2005) (discussing the challenge in the transnational context).

104. See Eric Goldman, *Content Moderation Remedies*, 28 MICH. TECH. L. REV. 1 (2021), <https://ssrn.com/abstract=3810580> or <http://dx.doi.org/10.2139/ssrn.3810580> (discussing nearly three dozen moderation actions taken by companies).

105. See *infra* text accompanying notes 130–139.

independent application to the instance of the matched image. Moreover, content moderation regimes might involve the execution of a single subfunction multiple times—often assigning the iterative performances to different actors (whether technical and human, or to different groups of humans)—such as through appeals or review processes that revisit rule-application or case resolution decisions.¹⁰⁶

b) Examining the Values Implicated, and Competencies Required,
by the Different Subfunctions

Deconstructing content moderation through this framework exposes the public values at stake in each, and the different competencies required for legitimate performance. These values and competencies are particularly important because platforms are asked to engage in governance activities that support the “public” or “common” interest¹⁰⁷ as distinct from their business interest.¹⁰⁸ To meet this goal, it is essential to ensure that the allocation of subfunctions align both with the competencies that are critical to procedural aspects of public interest regulation, such as independence and representativeness, and with the expertise and operational capacity necessary for actions and outcomes substantively aligned with the public interest.¹⁰⁹

Conversations about mixed public-private governance regimes typically focus on questions about the appropriateness of delegating aspects of implementation, or sometimes adjudication, to the private sector. Public law does not usually outsource the core role of policy formation or rulemaking to private entities.¹¹⁰ Agenda setting and guidance, including establishing

106. See Jean Patja Howell, *The Lawfare Podcast: How Zoom Thinks About Content Moderation*, LAWFARE (Dec. 2, 2021), <https://www.lawfareblog.com/lawfare-podcast-how-zoom-thinks-about-content-moderation> (podcast in which Zoom executives discuss the multiple layers at various stages of content moderation review conducted by different groups).

107. See Walter Mattli & Ngaire Woods, *In Whose Benefit? Explaining Regulatory Change in Global Politics*, in *THE POLITICS OF GLOBAL REGULATION* 1, 4 (Walter Mattli & Ngaire Woods eds., 2009) (distinguishing between “captured regulation” which entrenches narrow interests and “common interest regulation” that fulfills broader public purposes). Mattli and Woods argue that public interest regulation can emerge where “open forums, proper due process, multiple access points and oversight mechanisms” exist along with robust societal demand for action. *Id.* at 17.

108. See Elkin-Koren & Perel, *supra* note 11, at 858 (distinguishing between “public” functions related to the identification and removal of unlawful content from “private” functions driven by business-related content moderation imperatives).

109. Abbott & Duncan, *supra* note 90, at 3.

110. The Computer Fraud and Abuse Act (CFAA) provides a telling example of the problems that arise where the law makes such a delegation. 18 U.S.C. § 1030. Efforts to bring claims against individuals—researchers, users and employees—who’ve violated corporate terms of services, which set all sorts of limits on how individuals can interact with systems,

decisional criteria, fall generally within the purview of the public sector with private actors implementing them, exercising only limited discretion. Yet, given the constitutional, jurisdictional, and practical limitations on government efforts to regulate speech that many platforms and many members of society want moderated, platforms currently are doing exactly this core policymaking work. Importantly, even when governments could do it, they have abdicated their role at times, at least in part leaving the core definitional subfunction in private hands.

“Defining” work sets policy and, therefore, particularly around expression, implicates core public values. Substantively, public policy must protect individual rights and balance competing interests. Procedurally, legitimate policy formation requires stakeholder consultation, public participation, expert deliberation, reasoned decision-making, and transparency throughout, including publication of the final policy adopted. Defining work requires competencies: independence (that is, not beholden to one party or having an interest in the outcome), representativeness of relevant stakeholders, and two forms of expertise—substantive (subject matter) and political (to negotiate across stakeholders).

The application of rules and policies implicates different values: the reasonable, consistent, and fair interpretation and application of rules. This work requires particular competencies: independence from parties and from outcome; normative expertise, especially in the global context in which content moderation is enacted; and representativeness. Issues about jurisdiction and juries of peers get at this competence and point to the difficulty of theorizing and operationalizing it in this context.

Constraints on who can perform the “identifying” subfunction, like standing requirements in litigation, limit the category of parties who can raise concerns about content. This may lead to underenforcement of rules, particularly if the remaining parties enabled to do the identifying work are less resourced. Explicit assignment of the “identifying” subfunction can act as a check on the co-option of platforms’ technical and human resources by more financially or politically powerful interest groups. In assigning the “identifying” subfunction, regulators should consider the entity’s operational capacity and

have been limited by court concerns with the substantive and procedural concerns of outsourcing the definition of what is illegal to the private sector. *See* *United States v. Drew*, 259 F.R.D. 449, 464 (C.D. Cal. 2009) (concluding that construing the CFAA to cover terms of service would render the statute void for vagueness); *see* *Van Buren v. United States*, 141 S. Ct. 1648, 1662 (2021) (narrowing construction of unauthorized access under the CFAA to cover obtaining information from particular areas within a computer to which an individual does not have access, but not to cover accessing available information for improper purposes).

the extent to which over or under enforcement is more aligned with public values. Certain allocations of the "locating" function may promote interests such as competition, which may be of importance independently or because of the potential to support freedom of expression.

Finally, resolving cases, like applying definitions to particular content, implicates the values of consistency, proportionality, and fairness with regard to penalties and other remedies. There may be more tolerance for variations in how distinct platforms resolve cases, rather than apply governing standards, pursuant to voluntarily adopted policies. Yet concerns with the consistency, proportionality, and fairness of the remedies imposed remain. For these reasons, resolving work, which assigns penalties and other remedies, requires independence from parties and from outcome, as well as normative expertise.

c) Constraints Intended to Protect Values and Enhance Competencies

In some instances, regulators recognize that an entity to whom they are allocating a subfunction is not fully equipped to perform it, or to perform it legitimately. To address foreseen deficits, regulations sometimes direct or constrain how a subfunction is performed. Constraints take two forms: process constraints and limitations on the actors used for implementation.

Process constraints impose procedures on the execution of a subfunction. These range from requirements for transparency or secrecy of rules, policies, or outcomes to detailed rules setting out how the subfunction must be implemented. Actor constraints establish a preference for a specific kind of actor or a limit on what type of party can perform the subfunction—human or technical. The case studies discussed below highlight ways in which such constraints are explicitly imposed as well as places where they are predictable, if not explicitly prescribed, outcome of regulatory design choices.

2. *Understanding Elements of the Functional Framework Through Case Studies*

Existing content moderation regimes provide a rich set of examples to explore the allocation of subfunctions and the use of constraints. Commentary on different regimes further highlights how particular arrangements of the content moderation function put public values more or less at risk.

a) Case Studies of Subfunctions

i. Defining: Section 230 and the DMCA

Section 230 of Communications Act¹¹¹ and the Digital Millennium Copyright Act¹¹² represent the ends of the spectrum with respect to the allocation of the defining subfunction. While Section 230 does not explicitly define content to be moderated, by shielding platforms from civil liability for removing or blocking content that they consider “obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable,” the law gives those private actors wide berth to define the content subject to moderation. The broad protection against liability is neither contingent on the content’s illegality—even content that is constitutionally protected may be removed at the platform’s pleasure—nor on the platform’s motives.

The law also encourages platforms to empower users in moderating content, by limiting civil liability for providing technology that helps users restrict access to objectionable material. As a formal matter, such technology leaves the definition of objectionable to the users themselves; the tools can be used to remove whatever content a user deems objectionable. Yet, in practice, the tools’ provision of categories of content for filtering constrain this definitional flexibility. Moreover, such tools, only allow users to configure the content to which they personally are exposed, rather than shape what is in circulation. Together, these two provisions largely give companies the latitude to define through written text, practices, and tools, the content that circulates on their platforms.

In contrast, federal law defines the content to be moderated under the DMCA. While the processes and responsibilities for action set out in the statute give copyright holders the ability to influence what content is removed, the definition of removable content is tethered to the legal definition of copyright infringement.¹¹³ Several provisions are designed to ensure that only infringing material is subject to moderation. There are *ex ante* checks on copyright holders’ claims,¹¹⁴ and there is recourse to courts to resolve disputes over whether content is infringing. Moreover, while platforms may decide to

111. 47 U.S.C. § 230.

112. 17 U.S.C. § 512.

113. 17 U.S.C. § 512(c)(3) (tethering the use of the takedown process to copyright law, and in particular, requiring the complaining party to have a good faith belief that use is infringing); *CoStar Grp., Inc. v. LoopNet, Inc.*, 373 F.3d 544, 552 (4th Cir. 2004) (indicating that DMCA safe harbor is irrelevant in determining what constitutes a *prima facie* case of copyright infringement).

114. 17 U.S.C. § 512(c)(3)(A)(v) (“Good faith” attestation); 17 U.S.C. § 512(f) (penalties for misrepresentation); 17 U.S.C. § 512(g) (counter-notice and putback).

reject a takedown request, nothing in the law requires or even invites them to do so. Indeed, the statutory framework protects platforms against liability for removing or limiting access to content based on a copyright holder's complaint or any other good-faith basis, regardless of whether the material is ultimately determined to be infringing or not. Individuals whose content is the subject of a takedown request can object, and if a dispute persists, either party may file a suit in federal court. This recourse to the courts maintains the role of public law and public institutions in defining the content to be moderated under the DMCA.¹¹⁵ While much of the action under the law occurs in the notice and takedown process outside the courtroom, giving rise to misuse,¹¹⁶ recourse to the courts, along with other regulatory design choices below, maintain the public role in defining moderated content.¹¹⁷

These different statutory allocations of definitional work shape stakeholders' perceptions of the legitimacy of the content moderation activities platforms take underneath them.¹¹⁸ For example, because § 230 enables

115. The role of the courts proved essential to maintaining the balance under copyright law between the rights holder's interest and the public's interest and, in particular, to establishing that before issuing a takedown notification a complaining party must consider whether the use of the material constitutes fair use because "fair use is 'authorized by the law.'" *Lenz v. Universal Music Corp.*, 815 F.3d 1145, 1153 (9th Cir. 2016) ("We conclude that . . . fair use is 'authorized by the law' and a copyright holder must consider the existence of fair use before sending a takedown notification under § 512(c)."). However, the consideration of fair use requires only good faith, meaning the complaining party can only be held liable if they "knowingly misrepresented" their own understanding of whether the use of the copyright constituted fair use. *Id.* at 1154.

116. Jennifer M. Urban, Brianna L. Schofield & Joe Karaganis, *Takedown in Two Worlds: An Empirical Analysis*, 64 J. COPYRIGHT SOC'Y USA 483, 514 (2017) (concluding based on empirical research that §512 is used to address "privacy, defamation, and other disputes" and to target "non-infringing material"); Jennifer M. Urban and Laura Quilter, *Efficient Process or Chilling Effects? Takedown Notices Under Section 512 of the Digital Millennium Copyright Act*, 22 Santa Clara Computer & High Tech. L.J. 621, 667 (2006) (finding substantive problems with 31% of § 512(c) and (d) notices reviewed); Daniel Seng, *Copyrighting Copywrongs: An Empirical Analysis of Errors with Automated DMCA Takedown Notices*, 37 SANTA CLARA HIGH TECH. L.J. 119 (2020).

117. The statute establishes safe harbors for internet service providers to protect them from liability for copyrighted works others make available through and on their systems. It does not preclude the parties to a dispute about the use of a copyrighted work from accessing the courts. The statute creates an additional remedy under § 512(f) for bad faith issuance of a takedown notice. *See Lenz v. Universal Music Corp.*, 572 F. Supp. 2d 1150, 1154–55 (N.D. Cal. 2008).

118. Elizabeth Dwoskin, *Trump is Suspended From Facebook for 2 Years and Can't Return Until Risk to Public Safety is Receded*, WASH. POST (June 4, 2021), <https://www.washingtonpost.com/technology/2021/06/03/trump-facebook-oversight-board/> (describing Facebook's response to a ruling from its Oversight Board's ruling on the appropriateness of banning former president Donald Trump as "an attempt to clarify Trump's penalty and make the procedures

platforms to engage in unbounded, undeclared, and unfettered moderation, users find the experience of being moderated mysterious, arbitrary, and at times, biased. In contrast, stakeholders have raised concerns about the copyright holders' desire to ignore the statutory requirement to consider fair use before filing a takedown notice.¹¹⁹ Moreover, concerns about covered content have not raised questions about the legitimacy of the platforms' actions. Thus, the distinct allocations of the definitional subfunction have contributed to different perceptions of the legitimacy of platforms' moderation activities.

b) Identifying: RTBF and CSAM

The “right to be forgotten” provision in the EU General Data Protection Regulation, which requires search engines to delist search results that violate an individual’s privacy interests,¹²⁰ and the suite of statutes, including the PROTECT Act,¹²¹ which shape how platforms moderate child sexual abuse material, allocate responsibility for identifying regulable content to different actors. In each regime, the law on the books does not explicitly enlist platforms in the work of identifying content. However, other aspects of the regulatory frameworks incentivize platforms to take on identification work.

The Article 29 Working Party’s guidance clarifies that the individual claiming the right to be forgotten bears the responsibility for identifying content for erasing, deleting, or delisting, at least in the first instance.¹²² The guidance documents clarify that the law does not require search engines to

of the powerful social network, which is used by 3.45 billion people globally on a monthly basis, appear less arbitrary and opaque to the public”).

119. See, e.g., India McKinney & Ernesto Falcon, *Electronic Frontier Foundation Memo to Incoming Biden Administration*, EFF (Jan. 21, 2021), https://www.eff.org/wp/eff-transition-memo-incoming-biden-administration#_Toc57064038 (“Section 512 strikes a balance between the interests of service providers, copyright owners, and Internet users—but the system is not perfect. It is too easy for copyright owners . . . to have speech taken down, which comes at a high price for free expression and the public interest. The problem of false and abusive takedown notices is widespread and well documented.”).

120. GDPR, art. 17, 2016 O.J. (L 119) 1, 43 (“Right to Erasure (“right to be forgotten”)” (providing that “[t]he data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay and the controller shall have the obligation to erase personal data without undue delay” where one several grounds applies).

121. Pub. L. 108–21, 117 Stat. 650 (2003) (codified at 18 U.S.C. § 2258A).

122. *Guidelines on the Implementation of the Court Of Justice of the European Union Judgment on “Google Spain and inc. v. Agencia Española de Protección de Datos (AEPD) and Mario Costeja González” C-131/12*, at 6 (Nov. 26, 2014), http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp225_en.pdf [hereinafter *Guidelines Implementing Google Spain Judgment*] (“The ruling does not oblige search engines to permanently carry out that assessment in relation to all the information they process, but only when they have to respond to data subjects’ requests for the exercise of their rights.”).

proactively identify personal data for removal but only “to respond to data subjects’ requests for the exercise of their rights.”¹²³

However, because the regulation provides no guidance on the format or content of notices, the specificity with which the content being requested for erasure or restriction (delisting) must be identified is uncertain. This uncertainty complicates the interactions between users, hosts, and search engines during the identification subfunction.¹²⁴ On one hand, relevant guidance from the Article 29 Working Party directs requesters to “identify the specific URLs.”¹²⁵ On the other, that same guidance, as well as guidance from member states’ data protection authorities, affirms the data subject’s right to make erasure requests as they see fit¹²⁶ and, in particular, to use methods beyond whatever standardized intake forms entities provide.¹²⁷ This combination creates some uncertainty with respect to whether platforms may be required, at the very least, to assist in identifying content covered by the regulation.

Moreover, the law allows an individual to request delisting, or other action by a search engine, without requesting erasure from the host. This complicates the distribution of responsibility for identifying—and the related task of locating—content covered by the law. If a host alters the URL at which content resides, a search engine could end up returning content it has attempted to delist from queries on the name of the data subject. In such situations, it is unclear whether there is any shift in responsibility for identifying or more narrowly re-locating the content subject to delisting. In addition, if the information subject to delisting was publicly shared online by a controller, the controller must take “reasonable” steps to inform other entities of the data

123. *Id.* at 6.

124. The Art 29 Working Group Guidance; guidance issued by the Information Commissioner’s Office in the U.K. states that Individuals can make a request for erasure verbally or in writing.

125. *Guidelines Implementing Google Spain Judgment, supra* note 122, at 7. The European Data Protection Board issued its own guidelines, *see* Guidelines 5/2019 on the criteria of the Right to be Forgotten in the search engines cases under the GDPR (part 1), (July 7, 2020), https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_201905_rtfsearchengines_afterpublicconsultation_en.pdf, however the Board does not address the content of erasure requests specifically.

126. *Guidelines Implementing Google Spain Judgment, supra* note 120, at 7 (noting that “national data protection laws provide for great flexibility . . . and offer data subjects the possibility of lodging their requests in a variety of ways” and so while it may be convenient for data subjects to use forms and procedures online services set up “it should not be the exclusive way for data subjects to exercise their rights”).

127. *Id.*

subject's request.¹²⁸ This may create a responsibility to assist in locating the content subject to delisting on other platforms.

In contrast, even though the statutory framework governing child sexual abuse material¹²⁹ allocates formal responsibility for identification outside the platforms, platforms have taken on some shared responsibility for the identifying subfunction.¹³⁰ Some platforms have developed their own databases¹³¹ of “hashes,” which are numeric “fingerprints” of previously identified images of child sexual abuse, against which they screen content on upload. Some companies claim to use machine learning classifiers to identify

128. GDPR, art. 17, recital 66, 2016 O.J. (L 119) 1, 13, 43.

129. These include general criminal statutes prohibiting the making, printing, publishing, distribution, reproduction, transportation, and possession of child pornography, 18 U.S.C. §§ 2251, 2251A, 2252, 2252A, 2252B, 2260; as well as the statute establishing the National Center for Missing and Exploited Children (NCMEC), an independent private agency funded by the U.S. Department of Justice's Office of Juvenile Justice and Delinquency Prevention tasked with assisting with the identification and recovery of missing and exploited children. NCMEC coordinates programs to locate missing children, provides technical assistance and training to law enforcement and other stakeholders, and provides information and assistance services. 42 U.S.C. § 5773(b)(1). While existing federal statutes refer to child pornography, we use the term child sexual abuse material which more accurately captures the content of the images which depict the rape, sexual abuse, and sexual exploitation of children and is the preferred term among experts. *See The EARN IT Act: Holding the Tech Industry Accountable in the Fight Against Online Child Sexual Exploitation: Hearing on S. 3398 Before the S. Comm. on the Judiciary*, 116th Cong. 2 n.1 (2020), <https://www.judiciary.senate.gov/imo/media/doc/Shehan%20Testimony.pdf> [hereinafter Testimony of John Shehan] (testimony of John Shehan, Vice-President, Exploited Children Division, National Center for Missing & Exploited Children).

130. Platforms are not required to proactively search out CSAM for removal, and responsibility for identifying CSAM is left to others: victims, platform users, law enforcement. Where platforms have actual knowledge of CSAM offenses, they are required to file reports with the NCMEC. *See The Prosecutorial Remedies and Other Tools to end the Exploitation of Children Today Act of 2003 (PROTECT Act)*, Pub. L. No. 108-21, 117 Stat. 650 (2003) (codified at 18 U.S.C. § 2258A). Penalties for knowing failure to report are \$150,000 for the first violation and \$300,000 for each subsequent violation. 18 U.S.C. § 2258A(e). The law places mandatory reporting requirements on providers of electronic communication service and remote computing services to the public.

131. *See* Testimony of John Shehan, *supra* note 127, at 3 (stating that “many technology companies . . . actively detect and remove child sexual exploitation content” and “go above and beyond the requirements of current law and look for innovative methods to address child sexual abuse material and implement sophisticated tools and technologies to identify this content online, report it to NCMEC, and get it quickly removed”); *United States v. Miller*, 982 F.3d 412, 419 (6th Cir. 2020) (describing Google's use of proprietary hashing technology to create hashes of confirmed child sexual abuse images, and scan customer files on upload for matches which a Google employee might view and confirm as child sexual abuse material or might just send an automated report with the file to NCMEC).

CSAM imagery that is not yet cached in hash databases.¹³² Machine learning algorithms are trained on known CSAM to identify statistical patterns which are then used to identify potential CSAM in the wild.¹³³

Many platforms rely on a shared hash database hosted by the same non-profit entity, the National Center for Missing and Exploited Children (NCMEC),¹³⁴ which employs a cross-platform technology called PhotoDNA to scan images. These scans happen generally prior to allowing them to be uploaded, against the NCMEC database of CSAM. According to the inventor of PhotoDNA, Hany Farid, more than 95% of the nearly 18 million reports in 2018 to NCMEC's CyberTipline, constituting over 45 million pieces of identified CSAM, were from photo DNA.¹³⁵

The NCMEC database process further allocates identification tasks between various human and technical actors. Platforms can choose whether or not to share the child sexual abuse material they find on their networks as part of the mandatory reports.¹³⁶ Human NCMEC analysts determine whether reported images are CSAM and, if so, add hashes of the images to the NCMEC database.¹³⁷ Thus, the company reports provide continuous source material

132. See, e.g., Kristie Canegallo, *Our Efforts to Fight Child Sexual Abuse Online*, GOOGLE BLOG (Feb. 24, 2021), <https://blog.google/technology/safety-security/our-efforts-fight-child-sexual-abuse-online/>.

133. Hany Farid, *Reining in Online Abuses*, 19 TECH. & INNOVATION 593, 595 (2018).

134. NCMEC is authorized to receive and review the CSAM. 18 U.S.C. § 2258A(a), (b)(4).

135. *Fostering a Healthier Internet to Protect Consumers: Joint Hearing Before the Subcomms. on Comm'n's & Tech. & Consumer Prot. & Commerce, of the House Comm. on Energy & Commerce*, 116th Cong. 2 (Oct. 16, 2019) (testimony of Hany Farid, Professor, University of California, Berkeley).

136. 18 U.S.C. § 2258A(b), (b)(4) (stating that reports “may, at the sole discretion of the provider, include . . . [a]ny visual depiction of apparent child pornography or other content relating to the incident such report is regarding”). From conversations with knowledgeable experts, the authors believe that the major platforms routinely share the images and videos they identify with NCMEC. However, we do not have a citation to support this claim, and we do not know whether other entities behave similarly. While the large platforms engage in active efforts to screen for CSAM, a 2020 NCMEC reported that only 1,400 of the approximately 7,000 electronic service providers who are statutorily required to report had voluntarily registered to report with NCMEC CyberTipline, and of those 1,400, only 169 had actually filed a report in 2019. Testimony of John Shehan, *supra* note 127, at 4.

137. *South Korean National and Hundreds of Others Charged Worldwide in the Takedown of the Largest Darknet Child Pornography Website, Which was Funded by Bitcoin*, U.S. DEPT. OF JUST. (Oct. 16, 2019), <https://www.justice.gov/opa/pr/south-korean-national-and-hundreds-of-others-charged-worldwide-takedown-largest-darknet-child> (reporting that previously unknown CSAM is analyzed by NCMEC for potential inclusion in the hash database). As of February 2014, NCMEC analysts only review files attached to a report if the report indicates that either the reporting company reviewed it or that it was publicly available. U.S. DEPT. OF JUST., INFORMATION PAPER FOR PROSECUTORS AND LAW ENFORCEMENT OFFICERS: CYBERTIPS AND SUPPRESSION: AVOIDING AND DEFENDING AGAINST FOURTH AMENDMENT CLAIMS 4

that can be analyzed and used to update the database of *identified* CSAM used by other providers to quickly *locate*, and then remove (the only acceptable form of *moderation* in this context) it from their platform. Similarly, material found on personal hard drives and other media, such as USB flash drives and other portable storage drives, during law enforcement investigations—often triggered by reports to the CyberTipline—is hashed and fed into the NCMEC database.¹³⁸

A complicated set of factors, rather than a regulatory mandate alone, drives the use of technology to identify CSAM. These factors include NCMEC publicizing and encouraging the use of PhotoDNA; Microsoft’s decision to donate the technology to NCMEC for free licensing to eligible entities; and pressure from governments, advocacy organizations, and individual victims and families. The statutory framework that requires service providers to report CSAM to NCMEC and eliminates liability for doing so facilitates the creation of the shared database.¹³⁹

Prior to the use of PhotoDNA and other technology by platforms to identify CSAM, the identification task was left to other actors.¹⁴⁰ Given the limitations of the technologies currently in use—hash databases locate

(2021). Prior to February 2014, NCMEC’s analysts would “determine if the material constitutes a violation of law.” U.S. GOV’T ACCOUNTABILITY OFF., GAO-03-272, COMBATING CHILD PORNOGRAPHY: FEDERAL AGENCIES COORDINATE LAW ENFORCEMENT EFFORTS, BUT AN OPPORTUNITY EXISTS FOR FURTHER ENHANCEMENT 9 (2002). This shift responded to a decision holding NCMEC to be a state actor where it opened and examined a file attached to a CyberTipline Report that the provider had not reviewed. *See* U.S. v. Keith, 980 F. Supp. 2d 33, 41–42 (D. Mass. 2013). Reviewing material that a provider has previously reviewed falls within the private search exception, whereas reviewing material that has not been reviewed by a human at the reporting entity may not. *See* U.S. v. Reddick, 900 F.3d 636 (5th Cir. 2018) (holding that identifying photos through a PhotoDNA match was a private search and that law enforcement human review of those files did not exceed the private search doctrine). The question of whether NCMEC is a governmental entity or state actor shapes the current review process. In *U.S. v. Ackerman*, 831 F.3d 1292, 1306–1307 (10th Cir. 2016), the court held that NCMEC was a government entity or a state actor and its review of the emails and attachments contained in a CyberTipline Report that not been examined by the provider violated the Fourth Amendment. Subsequent legislation has attempted to clarify that NCMEC is a private non-profit entity, but it is unclear whether it has done so. *See* CyberTipline Modernization Act of 2018, Pub. L. No. 115-395, 132 Stat. 5287 (Dec. 21, 2018).

138. Interview with Hany Farid, Professor, U.C. Berkeley Sch. Info, on file with authors, June 26, 2020.

139. 18 U.S.C. § 2258A (establishing reporting requirements and penalties for failure to report); *id.* § 2258B (limiting liability for required reporting).

140. Given that platforms are not required to screen content for violations of law, they generally rely on users—and others—to identify (and generally require them to simultaneously locate) content that violates both law and platform policies. They use a range of techniques to do this, ranging from reporting forms to flags.

previously identified CSAM, and predictive classifiers are imperfect—the torrent of CSAM, and the rise of end-to-end encrypted services, the public (victims, advocacy organizations, etc.) and law enforcement continue to play an important role in identifying CSAM.

The different allocations of the identifying subfunction under the RTBF and the CSAM frameworks have produced various critiques. While the CSAM regime implicitly allocates the task of identifying content to be moderated on other stakeholders, as described above, it does not require platforms to establish procedures to assist individuals in reporting CSAM. The reporting infrastructures for CSAM on large platforms vary and have been criticized for being difficult to find and use, for failing to support victims,¹⁴¹ and for failing to remove CSAM consistently. Reporting structures for CSAM are more difficult to find and use than those for copyright, for example.¹⁴² Platforms do not have specific processes or flags within general content reporting processes to report alleged CSAM material. Within major platforms, “in nearly all cases it was impossible to explicitly flag content as CSAM.”¹⁴³ The complexity and inconsistency between desktop and mobile versions of the same platforms across different platforms make identifying CSAM material for action difficult. In many instances, the reporting functions are insensitive to the nature of the crime and victims—for example requiring reports to come through platform accounts, requiring identifying information, and requiring information about the alleged perpetrator without clear guidance about how it will be used. Similarly, the lack of clarity about who and how content subject to action under the right to be forgotten should be identified skews incentives. In the CSAM context, the lack of standardized identification requirements and process, particularly when contrasted with the clear process set out under the DMCA, burdens victims who are relatively less powerful, have less capacity, and are at risk to harm that cannot be remedied through damage awards. The allocation of identification work without clear procedures, in both the CSAM and RTBF context, has created hurdles for victims and their allies, as well as platforms, and contributed to concerns about the substantive commitment of platforms to addressing the substantive harms at issue.

141. CANADIAN CTR. FOR CHILD PROT., REVIEWING CHILD SEXUAL ABUSE MATERIAL REPORTING FUNCTIONS ON POPULAR PLATFORMS 7 (2020) (reporting that survivors of CSAM generally characterize the reporting experience as “disheartening,” lengthy, and sometimes futile, and they report being challenged by moderators).

142. *Id.* (“[U]sers concerned with issues related to copyright infringement, [sic] almost universally have access to formal reporting tools and clear instructions for initiating a complaint,” none of which are available for CSAM victims.)

143. *Id.* (describing research findings with respect to Twitter, Facebook, Microsoft, Google, Snapchat, TikTok, Discord, Pornhub, and Xvideos).

The NCMEC hash database described above¹⁴⁴ presents an example of the way that assignment for identifying content as problematic, and actually locating it online, can be bifurcated. While the statutory framework does not specifically allocate responsibility for either identifying or locating CSAM, it facilitated the emergence of a shared identification infrastructure in two ways. First, the statutory framework creates a centralized collection of images by requiring electronic service providers to report alleged CSAM to the NCMEC CyberTipline and eliminating any risk of liability for sharing CSAM material.¹⁴⁵ Second, the law allows NCMEC to share hash values of collected CSAM (but not the images themselves) with other electronic service providers for the exclusive purpose of stopping the sexual exploitation of children.¹⁴⁶

After material is identified and found to meet the definition (*application* subfunction discussed below) of CSAM, a hash of the image is included in the database. Participating entities receive the hashes of all CSAM content in the NCMEC database—those they have contributed, and those others have contributed—to aid in locating matching content on their networks. In this framework, material that is a candidate for moderation is sent to NCMEC whose analysts determine whether it is CSAM (rule *application*), and if so, add it to the hash database.¹⁴⁷ The hashes are then used to locate that same image or video (or a slightly perturbed version of it) on many platforms to facilitate further removal, reporting, and subsequent investigations.

Discretizing the subfunctions of identifying and locating has several potential benefits. First, NCMEC staff and law enforcement, as described above, apply the rules by determining what material enters the database.¹⁴⁸ These entities may be perceived as more expert, and therefore their decisions about whether rules apply to specific material are more legitimate. Second, platforms, other electronic service providers, and scholars have noted that proactive screening for CSAM material is costly and reviewing such images is emotionally difficult for employees.¹⁴⁹ The shared database of known CSAM

144. *See supra* text accompanying notes 134-138.

145. 18 U.S.C. § 2258(B).

146. 18 U.S.C. § 2258(c)(a). NCMEC also shares reported information with law enforcement. 18 U.S.C. § 2258(C)(d) (allowing NCMEC to make reports, including images, available to law enforcement).

147. The NCMEC database contains a subset of CSAM which its creators decided was indisputably illegal: images of children under the age of 12, who are typically prepubescent, involved in an explicit sexual act. Hany Farid, *Reining in Online Abuses*, 19 *TECH. & INNOVATION* 593, 598 (2018).

148. *Supra* text accompanying notes 134-138.

149. U.S. GOV'T ACCOUNTABILITY OFF., GAO-03272, *COMBATING CHILD PORNOGRAPHY: FEDERAL AGENCIES COORDINATE LAW ENFORCEMENT EFFORTS, BUT AN OPPORTUNITY EXISTS FOR FURTHER ENHANCEMENT* 2002 [hereinafter *COMBATING CHILD*

enables all platforms to benefit from the prior identification and reporting work of peer platforms and NCMEC's application work. The automated screening tool reduces the costs and improves the pace of content moderation and reduces the human toll of identification work. Third, the task of identifying and locating can impose a disproportionate burden on smaller companies. Sharing the benefits of previous identification and application work, and easing the burden of locating through automation, can reduce the cost and labor associated with moderating CSAM.¹⁵⁰ This shared infrastructure for CSAM can help platforms behave in ways that other stakeholders view as legitimate and socially responsible.

The allocations of content moderation subfunctions in CSAM has raised substantive and procedural concerns around privacy and due process. The NCMEC database allows platforms to allocate some identification work to NCMEC while sharing the location work through the PhotoDNA software, which allows them to screen content against it. Yet this intertwining of functions, combined with questions about whether NCMEC is a government agent or actor, has raised constitutional concerns.¹⁵¹ Courts in the United States have reached different opinions about whether or not this structure creates a special relationship between NCMEC and law enforcement.¹⁵² Congress has attempted to eliminate these concerns, and NCMEC has altered the way it handles images to avoid constitutional privacy concerns more clearly.

c) Applying: Section 230 and the DMCA

Section 230 and the DMCA provide examples of different ways the task of applying the rules to content identified as potentially subject to moderation can be allocated. Section 230 sits at one end of the spectrum, the DMCA at

PORNOGRAPHY]; George W. Burruss, Thomas J. Holt, & April Wall-Parker, *The Hazards of Investigating Internet Crimes Against Children: Digital Evidence Handlers' Experiences with Vicarious Trauma and Coping Behaviors*, 43 AM. J. CRIM. JUST. 433 (2018); Kathryn C. Seigfried-Spellar, *Assessing the Psychological Well-Being and Coping Mechanisms of Law Enforcement Investigators vs. Digital Forensic Examiners of Child Pornography Investigations*, 33 J. POLICE & CRIM. PSYCH. 215 (2018).

150. DEPT. FOR DIGIT., CULTURE, MEDIA & SPORT & HOME OFF., ONLINE HARMS WHITE PAPER 56 (2019) ("Badly designed regulation can stifle innovation by giving an advantage to large companies that can handle compliance more easily. We are determined that this regulatory framework should provide strong protection for our citizens while avoiding placing an impossible burden on smaller companies."); COMBATING CHILD PORNOGRAPHY, *supra* note 147 (reporting on industry feedback reporting that cost is a barrier to developing and using technology to proactively identify CSAM).

151. Tyler O'Connell, *Two Models of The Fourth Amendment and Hashing to Investigate Child Sexual Abuse Material*, 53 U. PAC. L. REV. 293, 309–14 (2021).

152. *Id.*

the other. As described above, the law does not obligate companies to adopt or enforce any rules about content, and to the extent platforms do so, they are free to apply them however they like, leading many to view the statute as an invitation to abdicate good governance.¹⁵³ However, this view provides only a partial picture. While the law places no affirmative obligations to moderate content, the drafters' goal from inception was to remove legal obstacles to platform content moderation.¹⁵⁴

By shielding platforms from liability for activities that would create liability as a publisher or secondary publisher (distributor) in the offline world, the law empowers platforms to engage in establishing rules and applying them to remove, edit, or otherwise moderate content without incurring any liability. Section 230's restraint on civil liability protects behind the scenes work that limit what content is available online and the development of settings and tools which users can use to moderate content. In both instances, the platform is determining the content to which its rules apply.

On the other end of the spectrum sits the DMCA. Unlike Section 230, which lumps entities who engage in a wide range of distinct functions into the single category of interactive computer service providers, the DMCA delineates different types of service activity in relation to third party content—hosting, locating, caching, transmitting—and specifies whether and what sort of actions entities must take to avail themselves of the safe harbor protections. For example, the statute directs requests for the removal of content to platforms that host content rather than search engines or caching services.¹⁵⁵

More importantly, while the law requires one of these functionally defined entities (hosts) to serve as a messenger between parties contesting whether the

153. See, e.g., Dawn C. Nunziato, *The Death of the Public Forum in Cyberspace*, 20 BERKELEY TECH. L.J. 115, 118 (2005) (arguing that “the government’s abdication of control over Internet speech regulation may well result in the loss of protection for speech that is insufficiently protected within an unregulated market for speech”); Danielle Keats Citron & Helen Norton, *Intermediaries and Hate Speech: Fostering Digital Citizenship for Our Information Age*, 91 B.U. L. REV. 1435, 1456 (2011) (noting that § 230 allows companies to decide whether and how to shape online expression and that while many opt to govern online hate speech, many others have not and some have built businesses around tolerating or encouraging online hate speech).

154. See JEFF KOSSEFF, *THE TWENTY-SIX WORDS THAT CREATED THE INTERNET* 63–64 (2019); see also *id.* at 57–76 (discussing drafters’ goals more generally).

155. DMCA, 17 U.S.C. § 512(b)(2)(E) (requiring caches to remove material only if it has previously been removed from the originating site or access to it has been disabled, or a court has ordered that the material be removed from the originating site or that access to the material on the originating site be disabled; and the party giving the notification includes in the notification a statement confirming that the material has been removed from the originating site or access to it has been disabled or that a court has ordered that the material be removed from the originating site or that access to the material on the originating site be disabled).

use of a copyrighted work is infringing, it leaves the application of copyright rules to that content to the parties or, if the parties choose, the courts. During the dispute between the parties, moreover, the statute explicitly controls how social media platforms, and others who host user generated content, should treat content—directing them to take it down expeditiously upon the receipt of a notice, restore it if a counter notice is received and the copyright holder does not notify the platform that they are seeking a court order for removal, and specifying timelines for the later actions.¹⁵⁶ The law shields social media platforms from liability for removing or limiting access to content based on a copyright holder’s complaint or any other good faith basis regardless of whether the material is ultimately determined to be infringing as long as they follow the safe harbor rules.¹⁵⁷ Both parties to a dispute have recourse to federal courts to resolve the claimed infringement, maintaining, to some extent, the role of public law in the application work of content moderation. Critiques of allocations of the application function provide useful insight into how constraints can address concerns with substantive and procedural legitimacy. Advocates and scholars have criticized platform moderation protected by § 230 for being inconsistent and opaque. Procedural critiques include the lack of transparency about rules and their application and the lack of an appeals process. Criticism has also focused on the actors platforms use to apply rules. Human actors tasked with moderating content may over- and under- block because as Sarah Roberts details in her research, both the leads and the front-line workers are often geographically and culturally removed from the platform they are monitoring.¹⁵⁸ Given this geographic and cultural removal, platform workers may lack the tacit knowledge necessary to fairly apply definitions to specific content.¹⁵⁹ Technical tools (in handoff parlance actors) are poor at accounting for information about culture, context, or use

156. *Id.*

157. 17 U.S.C. § 512(g)(1).

158. SARAH T. ROBERTS, BEHIND THE SCREEN 35 (2019) (“Both the headquarter from the tech platform and its audience may be very far removed, geographically and culturally, from the location where workers are viewing and moderating the user-generated content.”).

159. Given the lack of visibility into content moderation practices it is difficult to assess the extent to which various factors, including geographically and culturally diversity, affect content moderation decisions. One recent qualitative study documents sensitivity to cultural differences relevant to content moderation, difficulties in reconciling, and training efforts. *See* SABRINA AHMAD, “IT’S JUST THE JOB”: INVESTIGATING THE INFLUENCE OF CULTURE IN INDIA’S COMMERCIAL CONTENT MODERATION INDUSTRY 21 (2019) (finding Indian content moderators engage in a “holistic and critical *analysis* of this content for accuracy and legitimacy”).

that in many cases are essential to determining whether an item of content meets the definition of content to be moderated.¹⁶⁰

d) Resolving: § 230 and the DMCA

As with rule application, Section 230 and the DMCA also illustrate two ends of the spectrum with respect to case resolution. Under the DMCA, the only resolution actions are removal and restoration, and the statutory framework tightly controls the timing and other requirements of both. There is no discretion left to platforms and therefore a limited critique of those resolutions. In practice, however, some platforms offer copyright holders other options for resolving disputes. For example, when a video on YouTube is surfaced through the Content ID program, which features a database of copyrighted works submitted by owners, copyright holders may choose monetization over removal. When videos are uploaded to YouTube, they are scanned against these files, and when a match occurs, the video may be blocked or the owner may instead choose to run advertisements before the video plays and reap the revenue. Through Content ID, copyright holders can further leave the video on YouTube but restrict which applications or websites can embed it. The safe harbor provisions explicitly limit the moderation activities of entities that allow users to transmit and locate information and those who cache information—for example entities that allow users to transmit information may not modify the content or choose recipients. The resolution activities of hosts are also shaped by statutory provisions that mirror general contributory and vicarious liability standards.¹⁶¹

Section 230, in contrast, leaves platforms free to determine the range of appropriate resolutions. They are protected from liability for any resolutions—

160. Emma Llansó and her co-authors similarly distinguish between the use and risks posed by (1) artificial intelligence of various sorts used for proactive detection (what we call identification) of content potentially subject to moderation on the one hand and (2) automated evaluation and enforcement (what we call application and resolution) of the policies governing identified content on the other. They identify risks across both subfunctions such as false positives and false negatives; biased and/or discriminatory performance; escalating need for private data; inadequate human involvement and oversight; as well as specific risks of automated enforcement (application and resolution) including normalizing prior restraints on speech, the lack of due process, limits on human oversight, and limits on redress and accountability. *See* Llansó, *supra* note 3, at 9–10.

161. 17 U.S.C. §§ 512(c)(1)(A), (B). Abiding by the provision of the safe harbors has protected platforms from liability for contributory and vicarious liability. However, specific kinds of advertising and customer support has supported successful claims against platforms under inducement. *See* Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd., 545 U.S. 913, 936–37 (2005) (finding liability for inducing infringement where a platform marketed itself as a venue for sharing copyrighted material, explicitly encouraged users to upload copyrighted materials, and actively assisted them in doing so).

removing or otherwise limiting access to content they deem objectionable—taken in good faith, as well as any resolutions that information content providers or users make through technical tools platforms provide.¹⁶² As Jeff Kosseff explains in his book on the history of Section 230, “. . . companies will not be considered to be the speakers or publishers of third-party content, and they will not lose that protection only because they delete objectionable posts or otherwise exercise good-faith efforts to moderate user content.”¹⁶³ Courts have held that the “otherwise objectionable” and “good faith” language largely protect platforms’ discretion.¹⁶⁴

C. LESSONS FROM THE CASE STUDIES: THE TYPES OF CONSTRAINTS USED IN STRUCTURING SUBFUNCTIONS

These legal frameworks also illustrate the types of constraints used to structure the performance of different subfunctions in ways that promote relevant values and bring relevant competencies to bear. These include constraints on the process used in decision-making and constraints on the allocation of functions between particular technical and human actors.

a) Process Constraints

Sometimes statutory frameworks anticipate the ways in which allocations of subfunctions may put the rights and interests of other parties, or the public interest, at risk. They accordingly establish some combination of procedural

162. 47 U.S.C. § 230(c)(2) (shielding interactive computer services and users from liability for removing or limiting access to content or providing technical tools to help information content providers or others to restrict access to content); *see, e.g.*, *E360insight, LLC v. Comcast Corp.*, 546 F. Supp. 2d 605, 607–08 (N.D. Ill. 2008) (protecting Comcast against Internet marketing company’s tort claims arising from Comcast’s use of filters to block unsolicited emails).

163. KOSSEFF, *supra* note 154, at 65–66.

164. *Barnes v. Yahoo!, Inc.*, 570 F.3d 1096, 1105 (9th Cir. 2009) (barring a claim against Yahoo! for failing to remove nude photos after promising to do so); *Domen v. Vimeo, Inc.*, 433 F. Supp. 3d 592, 603–04 (S.D.N.Y. 2020) (stating that § 230(c)(2) “does not require that the material actually be objectionable; rather, it affords protection for blocking material ‘that the provider or user considers to be’ objectionable”); *Enigma Software Grp. USA, LLC v. Malwarebytes, Inc.*, 946 F.3d 1040, 1051–52 (9th Cir. 2019) (concluding that “if a provider’s basis for objecting to and seeking to block materials is because those materials benefit a competitor” their action would fall outside (c)(2) but that “the catchall [“otherwise objectionable”] was more likely intended to encapsulate forms of unwanted online content that Congress could not identify in the 1990s” and therefore provided interactive computer services with the ability to moderate content deemed objectionable beyond the “seven specific categories that precede” the term.) For an in-depth discussion of § 230(c)(2) case law, see Edward Lee, *Moderating Content Moderation: A Framework for Nonpartisanship in Online Governance*, 70 AM. U. L. REV. 913, 971–81 (2020).

constraints or limitations on the use or kind of automation or human actors used to implement the subfunction. The DMCA, for example, sets procedural constraints on many subfunctions of content moderation.¹⁶⁵ The law establishes relatively rigorous and balanced procedural rules that constrain the identification, application, and resolution subfunctions. The statute dictates the information that copyright holders must provide to initiate the takedown process (identification). It requires platforms and other hosts to designate an agent¹⁶⁶ to receive notices of alleged infringement. It also requires platforms to be transparent about the actions they take under the law, including notifying the party who posted the content that it has been removed, sharing the takedown notice that triggered removal, and informing them of their right to file a counter notice challenging the allegations in the takedown. It controls the timing of various aspects of the process, requiring platforms to “expeditiously” remove content when they receive such a notice¹⁶⁷ and to reinstate the content in “not less than 10, nor more than 14” days if the user files a counternotification and the complaining party does not notify the platform that they are filing a court action in response.¹⁶⁸ While the platform must expeditiously take down the content alleged to be an infringement, they must also promptly notify the party who posted the content to maintain protection against claims flowing from that removal. The subfunctions delegated to various parties are constrained through procedural protections aimed at protecting the interests of all parties to the dispute and, in particular, ensure that the opposing parties have access to basic information necessary to go to court. The procedural requirements thus ease the burdens placed on platforms, ensure that copyright holders have access to predictable and standardized complaint processes across platforms, and assure that users

165. See text accompanying at *supra* notes 112–118.

166. 17 U.S.C. § 512(c)(2) (stating the person’s contact information must be available on the website and from the Copyright Office).

167. 17 U.S.C. § 512(c) (stating they must meet a set of other criteria to be eligible for the safe harbor including they do “not have actual knowledge” of a copyright infringement).

168. 17 U.S.C. § 512(g)(2)(b).

receive information in a timely manner so they can assert their rights and are protected from frivolous,¹⁶⁹ overbroad,¹⁷⁰ and false claims.¹⁷¹

Procedural constraints come in many forms. They may be designed to ease the work placed on platforms or to protect the rights of competing individuals by regularizing processes and information flows, or they may tilt the field in ways that favor the interests of specific parties.

b) Constraints on the Allocation of Functions Between Particular Technical and Human Actors

Some content governance regimes require or prohibit certain kinds of actors from undertaking specific subfunctions. This may be affected through a requirement or incentive to use a particular protocol or database or a prohibition on automating certain subfunctions, at least when the results have particular effects on individuals.

169. 17 U.S.C. § 512(c)(3). They must be in written form and signed by “a person authorized to act on behalf of the owner” of the copyright; and contain a statement that the information is accurate, and under penalty of perjury, that the complaining party is authorized to act on behalf of the owner of an exclusive right that is allegedly infringed.

170. 17 U.S.C. § 512(c)(3)(A)(v) (providing that a valid DMCA complaint must include a “statement that the complaining party has a good faith belief that use of the material in the manner complained of is not authorized by the copyright owner, its agent, or the law.”). The Ninth Circuit applies a subjective standard for “good faith,” meaning “a copyright owner cannot be liable simply because an unknowing mistake is made,” even if the mistake was unreasonable. *Rossi v. Motion Picture Ass’n of America, Inc.*, 391 F.3d 1000, 1005 (9th Cir. 2004). Instead, “there must be a demonstration of some actual knowledge of misrepresentation on the part of the copyright owner.” *Id.* The Ninth Circuit has also held that the complaining party must consider whether the use of the material constitutes fair use before issuing a takedown notification. *Lenz v. Universal Music Corp.*, 815 F.3d 1145, 1153 (9th Cir. 2016) (“We conclude that . . . fair use is ‘authorized by the law’ and a copyright holder must consider the existence of fair use before sending a takedown notification under § 512(c).”). However, the consideration of fair use falls under the umbrella of good faith, so the Ninth Circuit also applies a subjective standard; the complaining party can only be held liable if they “knowingly misrepresented” their own understanding of whether the use of the copyright constituted fair use. *Id.* at 1154.

171. 17 U.S.C. § 512(f). In the case of a knowingly mistaken complaint, the DMCA creates a legal cause of action for the recipient of a false complaint to obtain a remedy from the issuer of the complaint. Because of the subjective standard that courts apply, a plaintiff bringing an action under § 512(f) must demonstrate that the complaining party knew they were issuing a false takedown notice. The Ninth Circuit allows the plaintiff to show “willful blindness” as a means of demonstrating subjective misrepresentation. *Lenz*, 815 F.3d at 1155. To prove willful blindness, a plaintiff must satisfy two factors: first, that the defendant subjectively believes “that there is a high probability that a fact exists,” and second, “the defendant must take deliberate actions to avoid learning of that fact.” *Id.* (citing *Global-Tech Appliances, Inc. v. SEB S.A.*, 563 U.S. 754 (2011)). In a § 512 (f) action, the plaintiff must show that the defendant knew there was a high probability that the content’s use was authorized and that the defendant deliberately avoided learning about its authorization. *Id.*

Rather than simply delegating responsibility to an entity, such constraints prescribe the allocation of workflows in an assigned subfunction between technical and human actors. For example, although § 230 does not limit or direct covered entities to use technology or rely on human judgment in specific ways, Congress intended to spur the market-driven development of filtering tools along with self-regulatory policies to address objectionable content.¹⁷² The law's stated objectives include "encouraging the development of technologies which maximize user control over what information is received by individuals, families, and schools" and "remov[ing] disincentives for the development and utilization of blocking and filtering technologies that empower parents to restrict their children's access to objectionable or inappropriate online material."¹⁷³ While it was drafted in part to spur the development, deployment, and adoption of "user empowerment" and "blocking and filtering" technologies¹⁷⁴ and to limit the burdens of human review that attach to publishing in traditional media, it does not explicitly require or constrain automation. To the extent that platforms provide technology to content creators to help them restrict access to material, the law limits their civil liability for providing it. These provisions, like the limitations on liability for third-party speech, were designed to incentivize companies to support technical approaches to content moderation.¹⁷⁵

Together, these provisions of Section 230 sought to spur the development of shared and proprietary content moderation infrastructures. At the time of its passage there was immense effort to develop technical standards,¹⁷⁶ rating systems, stand-alone products, and built-in "parental controls" to support

172. 141 CONG. REC. H8460 (daily ed. Aug. 4, 1995) (statement of Rep. Cox). For background on the purpose and history of the law, see Brief of Law Professors with expertise in Internet Law, as Amici Curiae Supporting Reversal, *Stephen J. Barrett M.D. v. Ilena Rosenthal* ¶ 4–11, 51 Cal. Rptr 3d 55 (2006) (No. S122953).

173. 47 U.S.C. § 230(b)(3)–(4).

174. *Id.* Purposes of the act include "to encourage the development of technologies which maximize user control over what information is received by individuals, families, and schools who use the Internet and other interactive computer services" and "to remove disincentives for the development and utilization of blocking and filtering technologies that empower parents to restrict their children's access to objectionable or inappropriate online material . . ." *Id.*

175. 47 U.S.C. § 230(b)(3) (establishing that "the development of technologies which maximize user control over what information is received by individuals, families, and schools" as a policy goal).

176. *Platform for Internet Content Selection*, W3C, <https://www.w3.org/PICS/> (last visited Mar. 24, 2022). The World Wide Web Consortium was developing the Platform for Internet Content Selection (PICS) which provided a specification to enable labels (metadata) to be associated with online content that rating services and filtering software could use. For a description of PICS, see *ACLU v. Reno*, 929 F. Supp. 824, 838–39 (E.D. Pa. 1996).

content moderation by end users. Some of these tools relied on content creators to identify their work as falling within a definition, while others relied on third parties to identify and label content, and some supported both.¹⁷⁷ The general approach regardless was to put decisions about what to moderate in the hands of end users. This goal was imperfectly met, as many tools provided very limited information to end users about how they defined and identified content.¹⁷⁸

Thus, Congress intended to provide individual users discretion over content to be moderated, and to allow companies to use and offer content moderation tools, to the extent the law allowed companies to shift power for defining, identifying, and moderating content to individuals, especially parents. Yet in practice the legislation shifted these tasks to third party tool providers. As with platform moderation policies and practices, these third party tools provided customers or users, as well as other stakeholders, with limited information about definitional criteria and implementation details, and were routinely found to engage in over blocking.¹⁷⁹ Content creators were often unaware that their material was being moderated, and if they objected, had no clear path to contest it.¹⁸⁰ While the drafters of the law understood that technology would play an important role in content moderation, they surely did not foresee the vast roles technology plays today or the unique opportunities and challenges it creates.

The DMCA also speaks to the use of technological actors. First, it requires platforms to “accommodate[s]” and “not interfere” with certain “standard technical measures” used to identify or protect copyrighted works.¹⁸¹ While it

177. For a description of the various filtering and rating technologies available at the time see *Am. C.L. Union v. Reno*, 929 F. Supp. 824, 839–42 (E.D. Pa. 1996), *aff’d*, 521 U.S. 844 (1997).

178. See *Fahrenheit 451.2: Is Cyberspace Burning?*, ACLU (Aug. 1997), <https://www.aclu.org/other/fahrenheit-4512-cyberspace-burning?redirect=fahrenheit-4512-cyberspace-burning> (providing a critique of the filtering technologies available at the time); Marjie Nouwen, Nassim Jafarainaimi, & Bieke Zaman, *Parental Controls: Reimagining Technologies for Parent-Child Interaction*, 15th PROC. EUROPEAN CONF. ON COMPUTER-SUPPORTED COOPERATIVE WORK—EXPLORATORY PAPERS, REP. OF THE EUR. SOC’Y FOR SOCIALLY EMBEDDED TECHS. 1, 2–3 (2017), <https://dl.eusset.eu/handle/20.500.12015/2928> (describing the four key functions of contemporary parental control technologies: time restrictions, content restrictions, activity restrictions, and monitoring and tracking).

179. For a collection of early studies and tests of Internet filters identifying over-blocking, see generally MARJORIE HEINS & CHRISTINA CHO, *INTERNET FILTERS: A PUBLIC POLICY REPORT IN FREE EXPRESSION POLICY PROJECT*, NATIONAL COALITION AGAINST CENSORSHIP (2001).

180. *Id.*

181. To qualify for safe harbors an entity must “accommodate[] and . . . not interfere with standard technical measures.” “[S]tandard technical measures’ means technical measures that

does not require copyright holders to use technical components to identify or protect their works, if they choose to, this non-interference requirement means hosting platforms must accommodate them on their infrastructure. Today, copyright holders routinely rely on technology to identify copyrighted works.¹⁸² Second, the law does not condition eligibility for the safe harbors on “monitoring. . . or affirmatively seeking facts indicating infringing activity,”¹⁸³ and case law has affirmed that platforms do not need to use technology to monitor or identify infringing material.¹⁸⁴ Thus while the law places a non-interference requirement on entities that host content, it does not require those same entities to independently invest in or use technology to monitor interaction with copyrighted material.

Finally, the GDPR as a whole contains a limit on completely automated decision-making. This provision could limit the capacity of platforms to use technical actors, or at least constrain how they are used, for specific subfunctions. Even where a decision is not based on a fully automated process or meets an exception to the prohibition on solely automated decision-making, entities must conduct an impact assessment of automated decision-making systems that pose a “high risk” to an individual’s rights and freedoms prior to adoption.¹⁸⁵ They must also provide individuals with explanations of the decisions such systems render. Guidance from the European Data Protection Board (EDPB)—an independent EU advisory body composed of representatives of the EU national data protection authorities and the European Data Protection Supervisor¹⁸⁶—on the GDPR principle of “privacy by design” emphasizes the need for evaluations of bias in algorithms that

are used by copyright owners to identify or protect copyrighted works and—have been developed pursuant to a broad consensus of copyright owners and service providers in an open, fair, voluntary, multi-industry standards process; are available to any person on reasonable and nondiscriminatory terms; and do not impose substantial costs on service providers or substantial burdens on their systems or networks.” 17 U.S.C. § 512(i).

182. Jennifer M. Urban, Joe Karaganis, & Brianna L. Schofield, *Notice and Takedown: Online Service Provider and Rightsholder Accounts of Everyday Practice*, 64 J. COPYRIGHT SOC’Y. USA 371, 374 (2017) (finding that automated “bots” are routinely used by large rights holders to search out infringements and generate takedown notices).

183. 17 U.S.C. § 512(m)(1).

184. *See* EMI Christian Music Grp., Inc. v. MP3tunes, LLC, 844 F.3d 79, 91 (2d Cir. 2016) (holding safe harbor protection could not be conditioned on service provider monitoring its service or affirmatively seeking facts indicating infringing activity); *Viacom Int’l, Inc. v. YouTube, Inc.*, 676 F.3d 19, 34 (2d Cir. 2012) (holding safe harbor protection could not be conditioned on affirmative monitoring by service provider).

185. *See* GDPR, art. 35, 2016 O.J. (L 119) 1, 53–54.

186. EDPB was created by the GDPR to replace the Article 29 Working Party. *See Article 29 Working Party*, EUR. DATA PROT. BD., https://edpb.europa.eu/about-edpb/more-about-edpb/article-29-working-party_en (last visited Mar. 18, 2022) [hereinafter *Article 29 Working Party*].

automate decision-making and the need for human oversight.¹⁸⁷ EDPB guidelines on targeting on social media impose specific requirements on the use of automated systems in certain contexts, affecting how technical actors are used in moderation.¹⁸⁸ Scholars have discussed the importance of these provisions in the context of automated content moderation.¹⁸⁹ More specifically, the provision of the Article 29 Working Party Guidance on the *Google Spain* decision which prohibits platforms from requiring individuals to funnel requests through particular forms or processes can be viewed as another constraint on the automation of the moderation process. Together, these provisions will influence the use of technology to enact some of the subfunctions of content moderation, although how and to what extent remains open.

Applying a functional framework to assess content moderation structures, then, involves asking three questions. (1) What subfunction is the structure intended to bolster or perform? (2) What are the public governance values implicated by that subfunction, and what is the set of competencies necessary to perform it? (3) What constraints, regarding the actor(s) used by the entity to which the subfunction is assigned, and the process they must use best protect relevant values and direct necessary competencies to the task? The next Part will consider the three symbolic legal structures discussed in Part II in light of

187. *Guidelines 4/2019 on Article 25 Data Protection by Design*, ¶ 70 (Apr. 2019), https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_201904_dataprotection_by_design_and_by_default_v2.0_en.pdf (explaining that privacy by design requires qualified human intervention capable of uncovering biases in automation in accordance with Article 22; that algorithms must be regularly assessed to assure they are fit for purpose and to identify and mitigate biases; and, that data subjects should be informed about the functioning of the processing of personal data based on algorithms that analyze or make predictions about them).

188. *Guidelines 8/2020 on the targeting of social media users*, ¶¶ 85, 86 (Aug. 2020), https://edpb.europa.eu/system/files/2021-04/edpb_guidelines_082020_on_the_targeting_of_social_media_users_en.pdf (stating that while prior guidance has found that “targeted advertising based on profiling will not have a similarly significant effect on individuals. . . . However, it is possible that it may do, depending upon the particular characteristics of the case, including: the intrusiveness of the profiling process, including the tracking of individuals across different websites, devices and services; the expectations and wishes of the individuals concerned; the way the advert is delivered; or using knowledge of the vulnerabilities of the data subjects targeted”); see also *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, at 22 (Aug. 22, 2018), <https://ec.europa.eu/newsroom/article29/items/612053/en>.

189. Llansó, *supra* note 3, at 13 (arguing that human oversight over automated decision-making in the context of content moderation is important at both the “individual decision level and through review of the systems that produced the error” because it “can provide a crucial safety net for the rights and freedoms of affected users”).

these questions and suggest ways that they might be used to structure content moderation going forwards.

IV. APPLYING THE FUNCTIONAL FRAMEWORK

The functional framework teases out (1) the values at stake in different subfunctions of content moderation; (2) the competencies required to ensure those values are protected when private actors govern speech in the online public sphere; and (3) the allocations and constraints aimed to address those competencies. This deeper understanding of the values at stake in privatizing different subfunctions provides a tool for assessing existing content moderation structures and constructing new ones.

In this vein, this Part applies the functional framework. It first assesses whether, or to what extent, the legally inspired structures adopted by platforms might either be empty symbols or meaningful efforts to address democratic subfunction deficits. Second, it illustrates how regulatory choices informed by the detailed understanding of the various ways subfunctions can be both allocated to leverage existing competencies and constrained to address deficits, and it illustrates how these choices can contribute to forward-looking decisions about content moderation processes that protect public values.

A. EVALUATING THE SYMBOLIC STRUCTURES

As described in Part II, platforms have fashioned a set of very visible structures in an attempt to backfill the democratic deficits created by particular allocations of subfunctions under current content moderation regimes. Google, through its Advisory Council, convened a body of diverse, independent experts to provide them with advice about *definitional* work based on their expertise, discussions with additional experts around the world, and public feedback. This diversified the expertise and expanded the demographics of the community shaping the important definitional work they were required to perform.

Facebook, through its oversight board, allowed greater scrutiny of the implementation of its rules in concrete cases—the *application* subfunction. By subjecting a subset of its content moderation activities to an independent review process it sought to associate its processes with the values of consistency, proportionality, rationality, and impartiality integral to the adjudicatory process.

Transparency reports have addressed the invisibility of some content moderation *resolution*. Creating a record of content removals and the legal bases behind them provides individuals and the public with an increased understanding of the policies, parties, and politics shaping the information

accessible online. Reports often provide detailed breakdowns on the kinds of removal requests,¹⁹⁰ on the countries from which removals are requested, and some information on the compliance with requests.¹⁹¹

The functional framework provides a starting point for more rigorous assessment of whether these symbolic structures, which surely nod towards these important public values, substantively address the democratic deficits associated with content moderation. This allows for analysis of the extent to which they actually build the requisite competencies to attend to the values at play in the discrete subfunctions.

190. See, e.g., *Content Removal Requests Report*, MICROSOFT CORP. SOC. RESP., https://www.microsoft.com/en-us/corporate-responsibility/content-removal-requests-report?activetab=pivot_1:primaryr3 (last visited Apr. 19, 2022) (stating that Microsoft breaks down its content removal into copyright, “right to be forgotten,” government requests (covered in their Content Removal Report), child sexual exploitation and abuse imagery, terrorist and violent extremist content, and non-consensual intimate imagery (covered in their Digital Safety Report); GOOGLE TRANSPARENCY REP., <https://transparencyreport.google.com/> (last visited Apr. 19, 2022) (showing that Google provides separate reports on de-listings under copyright, government requests, European privacy law, YouTube’s community guidelines, and the Network Enforcement Law); TWITTER TRANSPARENCY, <https://transparency.twitter.com/> (last visited Apr. 19, 2022) (showing that Twitter breaks content removals under their TOS out along many dimensions including violence, terrorism and violent extremism, child sexual exploitation, abuse and harassment, hateful conduct, and provides a separate report about removal requests which covers court orders, information identified by trusted reporters as illegal under local law, among others); *Transparency Reports*, META, <https://transparency.fb.com/data/> (last visited Apr. 19, 2022) (Facebook breaks down its reporting into three general categories, actions taken under their “community standards,” requests related to intellectual property, and “content restrictions” which includes removal requests from “governments and courts, as well from non-government entities such as members of the Facebook community and NGOs”).

191. Facebook, Google, Microsoft and Twitter each provide some information about requests within specific countries and compliance with them, but the detail provided varies. Facebook provides per country aggregates and some illustrative case studies. See *Restrictions by country*, META, <https://transparency.fb.com/data/content-restrictions/country> (last visited Mar. 24, 2022). Twitter provides per country reports. See *Japan*, TWITTER TRANSPARENCY, <https://transparency.twitter.com/en/reports/countries/jp.html> (last visited Mar. 24, 2022). For analysis, see *Removal Requests*, TWITTER TRANSPARENCY, <https://transparency.twitter.com/en/reports/removal-requests.html#2020-jul-dec> (last visited Mar. 24, 2022). Microsoft provides country-by-country breakdowns of requests. See MICROSOFT CORP. SOC. RESP., *supra* note 189. Google provides country-by-country breakdowns as well as notable examples. See *Government requests to remove content*, GOOGLE TRANSPARENCY REP., https://transparencyreport.google.com/government-removals/overview?removal_requests=group_by:requestors;period:&lu=removal_requests (last visited Mar. 24, 2022).

1. *Google Advisory Council*

The *Google Spain* decision and now the General Data Protection Regulation's Right to be Forgotten provisions leave platforms, including Google, with very little guidance on how to avoid liability. This regulatory configuration has the effect of implicitly allocating core definitional work—formulating rules about what content is subject to moderation—along with implementation responsibility (the application subfunction) to platforms.

Critics of Google's delisting system allege that Google serves as a private adjudicator, serving as "the judge and the jury." This is, of course, an outcome of the regulatory framework, which leaves both process and substance largely to Google's judgment.¹⁹² However, this critique doesn't capture concerns over the more unique and consequential delegation to private power: the policymaking, or definition-setting, function.

As Part II describes, the content moderation regimes emerging under the GDPR and the Digital Millennium Copyright Act sit at opposite ends of the spectrum when it comes to definitional work. The DMCA requires platforms to assist in implementation but the task of defining remains within the purview of public law. Platforms are not tasked with evaluating requests beyond their compliance with statutorily prescribed formalities or interpreting the law. In contrast, the GDPR gives platforms the initial responsibility for evaluating requests under a vague and ambiguous law.¹⁹³ While complaining parties can ultimately appeal to court if they dislike the platform's ruling, those whose content has been removed lack similar recourse.¹⁹⁴ This gives private platforms a much more central role in the content-governance process. As Edward Lee argues, in the right to be forgotten context, Google acts as a "private administrative agency exercising quasi-lawmaking, quasi-adjudicative, and quasi-enforcement powers."¹⁹⁵ Thus Google is both designing decision-making criteria under a broad legislative standard—akin to the detailed policy

192. Eldar Haber, *Privatization of the Judiciary*, 40 SEATTLE U. L. REV. 115, 137 (2016).

193. See Edward Lee, *Recognizing Rights in Real Time: The Role of Google in the EU Right to Be Forgotten*, 49 U.C. DAVIS L. REV. 1017, 1035 (2015) (asking "what institution should have the primary responsibility of addressing or clearing up those ambiguities?" and concluding that "Google has played a defining role in operationalizing the right to be forgotten and deciding what circumstances warrant a removal of a link to personal information or not" and that Google is delegated much authority).

194. Daphne Keller, *The Right Tools: Europe's Intermediary Liability Laws and the EU 2016 General Data Protection Regulation*, 33 BERKELEY TECH. L. J. 287, 359 (2018) (discussing GDPR art. 57(1)(f) which requires Data Protection Authorities to review claims based on data protection rights and concluding that a claimant cannot bring a free expression right challenge to a Right to be Forgotten action before them).

195. *Id.* at 1066.

work done by administrative agencies—and establishing the processes through which those rules are applied.¹⁹⁶ As Jean-Marie Chenou and Roxana Radu write, “the ‘right to be forgotten’ transforms a private intermediary into a quasi-legal decision-making body in charge of interpreting and implementing a law”¹⁹⁷

Google’s reliance on external experts and solicitation of public input in crafting policy is responsive to the tasks they were asked to undertake.¹⁹⁸ In forming the council and seeking advice through a broad consultative process, Google was backfilling to address the lack of clear procedural and substantive rules to guide their decisions.

Yet the implementation was fraught from the start. Some viewed the creation of the Advisory Council itself as an effort to undermine public deliberation and guidance. In particular, it was viewed by some as an effort to usurp the implementation guidance role played by the Article 29 Working Party,¹⁹⁹ which did eventually provide guidance in November 2014, a few months prior to the final report of the Advisory Committee. For example, Paul Nemitz, Director of Fundamental Rights in the European Commission, stated that Google could have “quickly, without fuss” implemented the ruling if they merely “outsourced [the takedown requests] to a normal midsize law firm in every member state” but chose instead to “start all this circus” which he viewed as “a very smart PR exercise.”²⁰⁰ Julia Powles argued that Google established the Advisory Council “to provide recommendations in parallel with, and in competition to, democratically legitimate regulators.”²⁰¹

Relatedly, Powles and others noted that Google’s charge to the Advisory Council and subsequent implementation constrained the deliberations of the

196. Christopher Kuner, *The Court of Justice of EU’s Judgment on the “Right to be Forgotten”: An International Perspective*, EJIL:TALK! (May 20, 2014), <https://www.ejiltalk.org/the-court-of-justice-of-eus-judgment-on-the-right-to-be-forgotten-an-international-perspective/> (stating that “[t]he judgment requires data controllers . . . to strike a ‘fair balance’ between these rights . . . but gives almost no criteria for doing so”).

197. Chenou & Radu, *supra* note 34, at 96.

198. See Lee, *supra* note 193, at 1071–72 (describing Advisory Council process and report as akin to a public agency’s rulemaking or recommendations and as a quasi-legislative function).

199. The Article 29 Working Party was an independent E.U. advisory body composed of all Member State data protection agencies. It was replaced, under the GDPR, with the EDPB, composed of representatives of the E.U. national data protection authorities and the European Data Protection Supervisor (EDPS). See *Article 29 Working Party*, *supra* note 186.

200. See Simon Davies, *Google’s “Right to be Forgotten” Offensive Goes Spectacularly Off the Rails*, THE PRIVACY SURGEON BLOG, (Sept. 27, 2014), <http://www.privacysurgeon.org/blog/incision/googles-right-to-be-forgotten-offensive-goes-spectacularly-off-the-rails/> (reporting public comments by Paul Nemitz),

201. Julia Powles, *The Case That Won’t Be Forgotten*, 47 LOY. U. CHI. L.J. 583, 591 (2015).

Advisory Council from the outset.²⁰² This context places the Advisory Council in a different light: rather than an effort to fill in work the government allocated to them, it suggests that the Advisory Council was in fact a move to usurp, or at least destabilize, some of the responsibility for defining work implicitly *allocated to a public authority*.

Regardless of the political posture, even in the best light some argued that the Advisory Council fell short on several dimensions essential to its legitimacy.

First, some claimed that the composition of the council lacked the robust stakeholder representation required of government advisory bodies. In particular, some argued that the data protection perspective—and in particular the perspective of the EU data protection authorities—was not well represented. Isabelle Falque-Pierrotin, the head of Article 29 Working Party, argued that the process constituted strategic theater, stating that Google “want[s] to be seen as being open and virtuous, but they handpicked the members of the council, will control who is in the audience, and what comes out of the meetings.”²⁰³ While the inclusion on the Council of José-Luis Piñar, the former Director of the Spanish Data Protection Agency and former Vice-Chairman of the European Group of Data Protection Commissioners, challenges this narrative to some extent, scholars reviewing the creation and activities of the Advisory Council found that the process sidelined public authorities. More generally, the Advisory Council members were viewed as having a speech-protective orientation, with few members of the Council supporting the *Google Spain* decision.²⁰⁴ Finally, some participants noted that the news coverage, which at times misconstrued the ruling, furthered the biases of the proceedings.²⁰⁵

Second, the Advisory Council had limited insight into the challenges faced by Google and other stakeholders. The charter of an advisory committee

202. *Id.* at 595–99 (discussing how Google shaped understanding of the ruling and the guidance provided by the Advisory Council).

203. Leila Abboud, *Google Hosts Meetings Across Europe on Privacy Rights*, REUTERS, (Sept. 8, 2014), <https://www.reuters.com/article/ctech-us-google-privacy-idCAKBN0H308I20140908>.

204. *See* Chenou & Radu, *supra* note 34 at 90–91 (describing perception that advisory council was dominated by individuals with a vested interest in a narrow construction of the right to be forgotten and, although not compensated for participation, other financial and policy entanglements with Google).

205. *See* Oral Testimony of Dr. Evan Harris, Trustee of Article 19, ADVISORY COMMITTEE PUBLIC MEETING, London, England (Oct. 16, 2014) (“[A]ny academic study of press articles would find probably a ratio of 9:1 opposition to the ruling, and to the rights that are supposedly due to be protected in that ruling. And that’s the right of the publishing world to do that, they have a vested interest.”).

supporting government activity generally establishes that it will receive the necessary support—both logistical and informational—to fulfill its mandate. Here, however, Google provided the Advisory Council with little information about its internal practices, leading a group of influential academics to write an open letter to Google about the process, demanding the release of data to inform the conversation.²⁰⁶ Pointing out the challenges of expert deliberation without access to critical information, Julia Powles wrote, “the expert hearings are largely conducted in a vacuum . . . leaving the debate to inapt analogies, rather than tales of real, human concern that inspire proactive responses.”²⁰⁷ Absent context, she concluded, the expert body insulated Google’s “processes with a veneer of authenticity and respectability” yet lacked real influence over them.²⁰⁸

Finally, the Advisory Council’s engagement with stakeholders and the public has been critiqued as relatively thin and performative.²⁰⁹ On the one hand, the creation of the Advisory Council and the seven public consultations in European capitals is a rather atypical effort at publicizing and formalizing efforts to bring in outside perspectives to inform firm policy. While platforms consult experts—including academics and civil society organizations—with some regularity, they typically do so behind closed doors and pursuant to non-disclosure agreements. The publicness of this consultation process, and its breadth, enabled a different level of stakeholder engagement. Numerous journalists, academics, and civil society thought leaders publicly commented on the proceedings, debates, and reports.²¹⁰

On the other hand, Julia Powles persuasively argued:

[I]hese hearings and their constrained format—where eight speakers were each given a short ten-minute window to present their high-level, often rather vehement, views—in practice, served as a

206. See Ellen P. Goodman, *Open Letter to Google From 80 Internet Scholars: Release RTBF Compliance Data*, ELLEN P. GOODMAN (May 14, 2015), <https://ellgood.medium.com/open-letter-to-google-from-80-internet-scholars-release-rtbf-compliance-data-cbfc6d59f1bd>.

207. Julia Powles, *Google’s Grand European Tour Seeks To Map Out The Future Of Data Ethics*, THE GUARDIAN (Sept. 10, 2014), <https://www.theguardian.com/technology/2014/sep/10/google-europe-explain-right-forgotten-eric-schmidt-article-29>.

208. Julia Powles, *How Google Determined Our Right to be Forgotten*, THE GUARDIAN (Feb. 18, 2015), <https://www.theguardian.com/technology/2015/feb/18/the-right-be-forgotten-google-search>.

209. See Chenou & Radu, *supra* note 34, at 92 (concluding that the Advisory Council “failed to engage public authorities and a wide range of viewpoints”).

210. For a compilation of academic commentary on the *Google Spain* decision, *supra* note 26, as well as the Advisory Council, see *Academic Commentary: Google Spain—Compiled by Julia Powles and Rebekah Larsen*, CAMBRIDGE CODE, <http://www.cambridge-code.org/googlespain.html> (last visited Mar. 24, 2022).

vehicle for individuals and organizations to express their discontent at various aspects of the ruling, as well as fermenting animosity to the ruling in press coverage.²¹¹

In short, the public engagement with experts did not facilitate a balanced exploration of views but rather generated more heat than light, and the public engagement with experts cast that heat in one direction. Finally, as described in Part II, although public comment was solicited, there is little evidence that it shaped the Advisory Council's deliberations. While additional experts and the public were provided some opportunity to speak, the process and final report provide little evidence that such participation and input meaningfully informed the deliberations.

2. *Facebook Oversight Board*

Mark Zuckerberg teased the public with the possibility of a board, made its creation a public spectacle, and has fashioned its activities to maximize its symbolism. Building an independent oversight board to review whether Facebook has accurately, fairly, and consistently applied its rules to particular content responds directly to missing competencies associated with the application subfunction—bringing in diverse, independent experts, creating transparent reviews, and building a body of precedent which creates predictability for stakeholders over time, thus regularizing and publicizing the application of rules. Yet, at every turn, the Facebook Oversight Board (FBOB) has been subject to criticism.

Some of that criticism manifests at the meta-level. Critics argue that Facebook chose a symbolic structure that focused attention on individual decisions—the application subfunction—to distract the public from the more important questions about the overall governing policies—the subfunction of *defining* the applicable rules. For example, the *Washington Post* Editorial Board called for Facebook to grant the Oversight Board greater ability to play an “advisory role” in initial content removal, rather than only coming in when a removal is appealed, since Facebook has faced the most controversy for the content it has refused to take down.²¹² Members of Congress pushed members

211. Powles, *supra* note 201, at 594.

212. Editorial Board, *Will Facebook's oversight board actually hold the company accountable?*, WASH. POST (May 17, 2020), https://www.washingtonpost.com/opinions/will-facebooks-oversight-board-actually-hold-the-company-accountable/2020/05/17/e1d46f50-93cd-11ea-9f5e-56d8239bf9ad_story.html (“For the time being, the board can only rule on material that has been removed from Facebook, not material that has remained on the site despite protestations [unless Facebook makes a referral]. Yet it is exactly these ‘leave-ups’ that catch the company the most flak.”). Kara Swisher went further arguing that “solving the problem of how to deal with speech across the largest and most unwieldy communications platform in

of the FBOB to demand authority to provide policy guidance to Facebook “to address the systemic amplification of divisive, racist, and conspiratorial content.”²¹³ They also asked the FBOB to obtain the power to publicly report metrics on their progress.²¹⁴ In addition, some felt this orientation drove attention away from legal reforms necessary to establish speech and privacy protection rules.²¹⁵ In addition, the timing of the FBOB’s creation was viewed as strategically designed to signal commitment yet forestall the adoption of important time sensitive, substantive protections.²¹⁶ Progressive advocacy organizations claimed that Facebook intentionally delayed the creation of the Board so as to avoid any board involvement in content decisions around the 2020 election.²¹⁷

Evaluated on its own terms as an effort to provide independent oversight and review of the *application* of Facebook’s rules to a limited set of cases subfunction, the FBOB built competencies that make it more than merely symbolic. First, several provisions protect the independence of the FBOB members, and while the initial set of members were hand-selected by Facebook,²¹⁸ future members will be selected without Facebook input. The

human history . . . may be beyond the capabilities of anyone,” noting that asking the oversight board to handle challenging content situations on a case-by-case basis “is trying to push back the ocean with one hand.” Kara Swisher, *Who’s Up for the Job of Decontaminating Facebook?*, N.Y. TIMES (May 6, 2020), <https://www.nytimes.com/2020/05/06/opinion/facebook-independent-oversight-board.html>.

213. Letter from the H. Comm. on Energy & Com. to Catalina Botero-Marino, Dean, Universidad de los Andes (Aug. 11, 2020), https://energycommerce.house.gov/sites/democrats.energycommerce.house.gov/files/documents/Botero-Marino.2020.8.11.%20Letter%20re%20Facebook%20Oversight%20Board.CAT_.pdf [hereinafter Letter to Catalina Botero-Marino] (“[W]e worry that your presence on the Board may help legitimize an entity that likely will have no ability to stop Facebook from amplifying conspiratorial and divisive content in search of advertising revenues.”).

214. *Id.*

215. Editorial Board, *supra* note 211 (“[T]he responsibility for addressing [privacy and content concerns] shouldn’t lie with the board, and it shouldn’t even lie with Facebook: Governments ought to set rules of their own, and [the United States] legislature has fallen short.”).

216. *Advocacy Groups Release Open Letter Urging Facebook Oversight Board Members To Take A Stand*, ACCOUNTABLE TECH (July 20, 2020), <https://accountabletech.org/media/press-release/> (complaining that the FBOB will not be operational until after the high-stakes 2020 elections).

217. *Id.* (reporting on the letter which also called on the five American members of the oversight board to step down).

218. See Klonick, *supra* note 53, at 2456–67 (discussing selection process, including close observation of potential members during table top simulations during consulting phase of FBOB development to determine whether they met criteria FB felt important, including open-minded, active listening, issue matter expertise, and dealing with adverse opinions); *Oversight Board Charter*, *supra* note 59, at art. 1, § 8 (describing initial set of 11 members who would then

independence of the initial set is protected by provisions that (1) allow removal only by the Trust established by Facebook to fund and oversee the Board; (2) allow removals only for a violation of policy rather than for the substance of decisions; and (3) provide for up to three, three-year terms.²¹⁹ The Trust was provided with only enough capital to support its operation through two three-year terms,²²⁰ raising questions of the Board as a whole's independence—would funding be renewed if Facebook does not approve of decisions?—and its long term viability. Members of Congress expressed concerns that the Board would not be sufficiently empowered to make independent, meaningful decisions²²¹ and that it might act as a “smokescreen,” allowing Facebook's executives to continue to make final decisions regarding content removal.²²² Members of Congress urged further assurances of board members' independence, asking them to commit to resigning if Facebook failed to grant them meaningful powers.²²³ Kate Klonick, who was provided special access to the process leading up to the FBOB creation, believes that the structure will allow the FBOB “to develop and maintain intellectual independence in the long-term.”²²⁴

While the charter constrains the decisions the FBOB can review, within its jurisdiction, many features protect the values at risk in the privatization of adjudication. The FBOB has the capacity to control its docket, an important aspect of independence, follows processes that provide opportunities for parties to be heard, and makes many of its actions transparent to the public.

select others over time to fill out the Board); Klonick, *supra* note 53, at 2461 (discussing divergence from charter, and ongoing involvement of Facebook in selection process).

219. *Oversight Board Charter*, *supra* note 59, at art. 1, § 8.

220. Elizabeth Culliford, Facebook pledges \$130 million to content oversight board, delays naming members, REUTERS (Dec. 12, 2019), *available at* (<https://www.reuters.com/article/us-facebook-oversight/facebook-pledges-130-million-to-content-oversight-board-delays-naming-members-idUSKBN1YG1ZG>) (describing Facebook's allocation of an irrevocable grant of \$130 million).

221. Maggie Miller, *Facebook Oversight Board to Address Racist, Voter Suppression Content*, THE HILL (Aug. 11, 2020, 07:10 PM EDT), <https://thehill.com/policy/technology/511591-house-democrats-pressure-facebook-oversight-board-to-address-racist-voter>; Letter to Catalina Botero-Marino, *supra* note 213.

222. Letter to Catalina Botero-Marino, *supra* note 213, at 2 (“[W]e worry that your presence on the Board may help legitimize an entity that likely will have no ability to stop Facebook from amplifying conspiratorial and divisive content in search of advertising revenues.”).

223. *Id.* at 4.

224. Klonick, *supra* note 55 at 2484 (concluding that Facebook's exclusion from the process, along with the disqualification of current and former employees, and the procedural affordances for the FBOB and Trust will assure its independence).

The Board has the discretion to accept and refuse reviews from all parties.²²⁵ The FBOB makes a subset of content removals more transparent to the public. Unlike the dry statistics provided by the transparency reports, the FBOB decisions provide a view into the nitty gritty of content moderation work done by Facebook itself.²²⁶ The FBOB is directed to publish and make decisions publicly accessible in a database and report statistics on the number and type of cases reviewed, cases submitted by different regions, and the timeliness of their reviews.²²⁷ While it was unclear under the bylaws whether requests from Facebook would be similarly published, and publicly archived, that is the FBOB practice. While covering a limited set of removals, the transparency is more useful in assessing how and why Facebook removes content. The FBOB decisions are binding on the particular piece of content at issue and establish precedent to guide future decisions.²²⁸ Facebook is required to implement decisions, unless doing so “could violate the law.”²²⁹

The limits on the FBOB’s jurisdiction are significant, however, and undermine the substantive value of the FBOB. While the FBOB is charged with “review[ing] content and issu[ing] reasoned, public decisions,”²³⁰ it “cannot review removals under local law,²³¹ and the FBOB reviews a minuscule number of the millions of instances in which Facebook applies rules to content.²³² The language of the charter would allow users to challenge both

225. *Oversight Board Charter*, *supra* note 59, at art. 2, § 1.

226. Facebook Oversight Board’s decisions provide detailed facts about actions Facebook has taken, including content removals, under corporate policies, and determines whether its actions accord with policy. *Board Decisions*, OVERSIGHT BOARD, <https://oversightboard.com/decision/> (last visited Mar. 24, 2022).

227. OVERSIGHT BOARD BYLAWS, art. 2, § 2.3.2, <https://www.oversightboard.com/sr/governance/bylaws/> (last visited Mar. 28, 2022).

228. *Oversight Board Charter*, *supra* note 59, at art. 2, § 2

229. *Id.* at art. 4.

230. *Id.* at art. 5 § 1.

231. *Id.* at art. 7.

232. In the third quarter of 2021 alone Facebook reports removing 13.6 million pieces of content for violating the violence and incitement policy and 9.2 million pieces for violating rules against bullying and harassment content. *Q3 2021 Community Standards Enforcement Report*, META (Nov. 9, 2021), <https://transparency.fb.com/data/community-standards-enforcement/?from=https%3A%2F%2Ftransparency.facebook.com%2Fcommunity-standards-enforcement>. In the second quarter of 2021 Facebook reported removing 20 million pieces of content from Facebook and Instagram globally for violating their policies on COVID-19-related misinformation. Guy Rosen, *Community Standards Enforcement Report, Second Quarter 2021*, META (Aug. 18, 2021), <https://about.fb.com/news/2021/08/community-standards-enforcement-report-q2-2021/>.

removal and non-removal decisions,²³³ but the bylaws narrow its jurisdiction by limiting reviews to single pieces of organic content that have been removed from Facebook and Instagram.²³⁴ The FBOB can only review a removal that has exhausted Facebook's internal review process.²³⁵ Finally, individuals must "have an active Facebook or Instagram account" to appeal to the Board.²³⁶ Facebook can ask the FBOB to review "questions related to the treatment of content beyond whether the content should be allowed or removed completely,"²³⁷ and the charter provides Facebook access to an expedited review process in exceptional circumstances.²³⁸

Together, these limitations create barriers to FBOB review of many consequential removals and non-removals and limit who can seek redress. While all courts have jurisdictional limits and standing requirements, the rationales for the limits set in the FBOB charter and bylaws—and in particular the narrowing between the two—have not been articulated or subject to discussion and debate. There is public concern with what Jack Balkin calls *collateral censorship*²³⁹—regulations that use private organizations to regulate the speech of others—and the ongoing tension between company commitments to international human rights norms and responding to local law. In this light, limiting FBOB's review of all actions taken under local law sidelines them in areas of profound importance, as the Board must direct its energy toward Facebook's private rules rather than engagement with public governance.

Finally, the FBOB leaves stakeholders in the position of largely taking whatever policies and processes the company offers. Consider the following:

The charter states that "[t]he purpose of the board is to protect free expression by making principled, independent decisions about important pieces of content and by issuing policy advisory opinions on Facebook's content policies."²⁴⁰ However, the emphasis on reviewing decisions to remove specific pieces of content limits the extent and manner to which questions of how those policies conform to freedom of expression, human rights norms

233. See *Oversight Board Charter*, *supra* note 59, at art. 2, § 1 ("[A] request for review can be submitted to the Board by either the original poster of the content or a person who previously submitted the content to Facebook for review"); Klonick, *supra* note 55, at 2463.

234. OVERSIGHT BOARD BYLAWS, *supra* note 227, at art. 3, § 1.

235. *Oversight Board Charter*, *supra* note 59, at art. 2, § 1.

236. OVERSIGHT BOARD BYLAWS, *supra* note 227, at § 1.2.2. Somewhat bizarrely, access to the Board is explicitly excluded for content and account holders in other Facebook services, including WhatsApp, Messenger, Instagram Direct, and Oculus. *Id.* § 1.2.1.

237. *Oversight Board Charter*, *supra* note 59, at art. 2, § 1.

238. *Id.* at art. 3, § 7.2.

239. Jack M. Balkin, *Old-School/New-School Speech Regulation*, 127 HARV. L. REV. 2296, 2298–99 (2014).

240. *Oversight Board Charter*, *supra* note 59, at 2 (Introduction).

generally, and other public values can arise. Like transparency reports, a key function of the FBOB is to channel public attention in particular ways—to focus on individual outcomes rather than fundamental questions about the appropriateness of Facebook’s policies and approaches to implementing them at scale. While the FBOB moves beyond the symbolic, providing protection for core values in the extremely limited and narrow set of cases it reviews, it channels public attention away from an arguably more important set of questions about the legitimate creation of rules—*defining* subfunction—and the entanglement of private platforms in government action. While independent review of platforms application of rules bolsters the protection for freedom of expression and other human rights, the limits on the FBOB’s jurisdictional scope, and the fraction of removals it can actually review,²⁴¹ limit its substantive value.

3. *Transparency Reports*

In their current form, transparency reports build operational capacity allowing companies to provide the public with improved understanding of the content removals that limit the information available on the web. However, they are partial and strategic, illuminating the impacts of a slice of the subfunction of *resolution*. They often provide information on a limited set of content removals—specifically the fact that content was removed pursuant to a particular law in a particular country—and provide limited data on that set. They do not reveal the identities of the actor requesting removal or the reasoning behind a removal decision. Thus, while they provide information about general trends and practices at a company over time, and in this way produce data similar to that of the wiretap reports they emulate, they fall far short of the records produced by court actions.

Some scholars view transparency reports as a “major step”²⁴² in providing “horizontal transparency”²⁴³—allowing stakeholders a glimpse into the corporate black-box.²⁴⁴ Yet even fans of the transparency reports note their “major limitations.”²⁴⁵ The quantitative emphasis of the reports leave stakeholders swimming in data but unable to discern its meaning. For example, in 2018 Facebook began including data about removals by their

241. See Klonick, *supra* note 55, at 2490 (reporting that based on 2019 figures that “approximately 170,000 pieces of content per day that would be potentially eligible for Board review”).

242. Gorwa & Ash, *supra* note 6, at 298.

243. Hans Krause Hansen & Mikkel Flyverbom, *The Politics of Transparency and the Calibration of Knowledge in the Digital Age*, 22 ORG. 872, 889 (2015).

244. Gorwa & Ash, *supra* note 6, at 292–94.

245. *Id.* at 302.

content review team under Facebook's Community Standards.²⁴⁶ Topics governed by the Community Standards include "adult nudity and sexual activity," "suicide and self-injury," "bullying and harassment," and "fake accounts" among others.²⁴⁷ The most recent Community Standards Enforcement Report states that 98.8% of "suicide and self-injury" content they took action on (including but not limited to removal) was identified by Facebook (referred to as proactively detected) and the remainder was reported by users.²⁴⁸ Given the nuanced and subjective judgments required to apply the suicide and self-injury policy that, for example, prohibits posting "content that focuses on depiction of ribs, collar bones, thigh gaps, hips, concave stomach, or protruding spine or scapula when shared together with terms associated with eating disorders" but allows those same images "in a recovery context," and similarly prohibits "content that depicts graphic self-injury imagery" but allows "older instances of self-harm such as healed cuts or other non-graphic self-injury imagery in a self-injury, suicide or recovery context,"²⁴⁹ it is hard to know what this statistic means beyond the fact that Facebook believes its employees and algorithms are well-tuned to their policy.²⁵⁰

The reports, moreover, make it difficult to interpret differences in performance across content and over time. For example, evaluating within and across policy areas and over time is complicated by varying definitions of

246. Sarah Perez, *Facebook's New Transparency Report Now Includes Data on Takedowns of 'Bad' Content, Including Hate Speech*, TECHCRUNCH (May 15, 2018), <https://techcrunch.com/2018/05/15/facebooks-new-transparency-report-now-includes-data-on-takedowns-of-bad-content-including-hate-speech/> (discussing a section of the report labeled the "Community Standards Enforcement Report").

247. *Community Standards Enforcement Report*, TRANSPARENCY CENTER, <https://transparency.fb.com/data/community-standards-enforcement/> (Last Visited May 10, 2022)

248. *Community Standards Enforcement Report: Suicide and Self-Injury*, TRANSPARENCY CENTER, <https://transparency.fb.com/data/community-standards-enforcement/suicide-and-self-injury/facebook/> (Last Visited May 10, 2022)

249. *Facebook Community Standards: Suicide and Self Injury Policy*, TRANSPARENCY CENTER, <https://transparency.fb.com/policies/community-standards/suicide-self-injury/> (Last visited May 10, 2022).

250. For example, did Facebook's identification of material stop posting—meaning that users had no opportunity to catch it? Did such identification rely on automated tools? Facebook reports that in Q4 2021 only 200 items of content actioned under the suicide and self-injury policy were appealed resulting in 50 restorations. They further report that they independently restored 95,200 pieces of content of their own accord. This provides limited information on how the policy is applied and what content it effected. *Facebook Community Standards: Suicide and Self Injury Policy*, TRANSPARENCY CENTER, <https://transparency.fb.com/policies/community-standards/suicide-self-injury/> (Last visited May 10, 2022).

“piece of content.”²⁵¹ More information is necessary to understand the meaning of these numbers and what effect they might have on the public.

The voluntary nature and lack of standardization undermine the substantive impact and utility of these symbolic structures.²⁵² Transparency reports share some common features—they generally provide quantitative information on legal requests received on a country-by-country basis, sometimes broken down by issue, and they often provide information on the proportion of requests with which the firm complied. Yet despite efforts to promote standardization,²⁵³ firms have not developed shared definitions or formats.²⁵⁴ When information is missing, it can be unclear whether the company deems it insignificant, is legally constrained from revealing it,²⁵⁵ or is intentionally withholding it.²⁵⁶

Transparency reports, moreover, are not merely partial; they are strategically so. Transparency reports “obfuscat[e] and redirect from more

251. *How Meta Improves: Content Actioned*, TRANSPARENCY CENTER, <https://transparency.fb.com/policies/improving/content-actioned-metric/> (last visited May 10, 2022).

252. For an overview of some of the variations in definitions and reporting categories and styles, see generally LIZ WOOLERY, RYAN H. BUDISH, LEVIN BANKSTON, *THE TRANSPARENCY REPORTING TOOLKIT: BEST PRACTICES FOR REPORTING ON US GOVERNMENT REQUESTS FOR USER INFORMATION* (2016).

253. *Id.*; *Transparency Reporting Guidelines, Release*, GOV'T. OF CANADA (June 30, 2015), <https://www.ic.gc.ca/eic/site/smt-gst.nsf/eng/sf11057.html>.

254. Independent organizations have stepped in to press companies to address the limitations of their transparency reports. Ranking Digital Rights publishes a yearly ranking of select companies' policies relating to freedom of expression and user privacy. Its index provides an important benchmark supporting comparative analysis across companies and over time. The initial effort to provide transparency around DMCA complaints, *chillingeffects.org*, is now *Lumen* and collects and analyzes a broader range of content removal requests from around the globe providing a source of independent data on content removals and their impacts. *LUMEN*, *supra* note 69.

255. See Losey, *supra* note 63, at 3454–55 (describing various jurisdictions' limits on disclosing law enforcement requests).

256. See, e.g., Rebecca Rose, *Google on Its Own Transparency Report: This is Not Good Enough*, *THE ATLANTIC* (Nov. 14, 2013), <https://www.theatlantic.com/technology/archive/2013/11/google-on-its-own-transparency-report-this-is-not-good-enough/281473/> (discussing Google's 2013 report, in which it noted that it had received more requests from the government than ever before, but could not share data about all of the government requests due to the Department of Justice's contention that requests related to national security must remain private); Sam Shear, *TikTok Transparency Report shows it removed 49 million videos and had 500 government requests*, *CNBC* (July 9, 2020), <https://www.cbc.com/2020/07/09/tiktok-transparency-report.html> (discussing TikTok's 2020 transparency report which had no data related to use of the app in China, where its parent company is based). The app goes by a different name in China, but it is unclear whether they intend to release information about that app in an additional report.

substantive and fundamental questions about the concentration of power, substantial policies and actions of technology behemoths.”²⁵⁷ The reports emphasize quantitative data, which is useful for some forms of analysis but lack the detail necessary to review and understand the legitimacy of legal demands and disclosures.²⁵⁸

Strategically partial reports limit more meaningful analysis. Statistics about the performance of machine learning tools, focused on accuracy and prediction, gloss over the human toll of content moderation failures. This emphasis on quantitative-transparency measures as the way to understand and evaluate content moderation skews public debates. All false labels are consequential but, in content moderation, some are far more so than others. For example, the recent failure of Facebook to remove the 3,000-member group calling themselves the Kenosha Guard that organized the armed response to the unrest in Kenosha, during which Kyle Rittenhouse shot three people, killing two, is a significant failure even if it was part of a quantitatively small number of pages that should have been removed.²⁵⁹

Thus, by producing statistics, rather than case studies or the more substantial records found in the Chilling Effects and Lumen database,²⁶⁰ these

257. Monika Zalnieriute, “Transparency Washing” in the Digital Age: A Corporate Agenda of Procedural Fetishism, 8.1 CRITICAL ANALYSIS OF L. 139, 139–53 (2021).

258. See, e.g., Losey, *supra* note 65, at 3454 (finding that Facebook, Google, LinkedIn, Microsoft, Tumblr, Twitter, Verizon, and Yahoo were not reporting the legal processes used by non-U.S. governments to access user data).

259. A Tuesday afternoon post read, “Any patriots willing to take up arms and defend our city tonight from the evil thugs?” Despite multiple reports under Facebook’s terms of service by users, the account was left up until after the shooting. There are multiple provisions of Facebook’s terms of service that could have led to the Kenosha Guard page being removed. Examples include provisions against Incitement to violence which bars “[s]tatements of intent or advocacy, calls to action, or aspirational or conditional statements to bring weapons to locations, including but not limited to places of worship, educational facilities or polling places (or encouraging others to do the same).” See *Violence and Incitement*, META https://www.facebook.com/communitystandards/credible_violence (last visited Mar. 24, 2022); see also *Dangerous Individuals and Organizations*, META, https://www.facebook.com/communitystandards/dangerous_individuals_organizations (last visited Apr. 20, 2022) (provisions limiting posts by “Dangerous Individuals and Organizations”); These provisions were updated on August 19, 2020 to more explicitly remove accounts hosting discussions of potential violence adopted to address militia organizations and others, *An Update to How We Address Movements and Organizations Tied to Violence*, META (Aug. 19, 2020), <https://about.fb.com/news/2020/08/addressing-movements-and-organizations-tied-to-violence/>, and eventually served as the basis for the group’s removal. See Russell Brandom, *Facebook takes down ‘call to arms’ event after two shot dead in Kenosha*, THE VERGE (Aug. 26, 2020), <https://www.theverge.com/2020/8/26/21402571/kenosha-guard-shooting-facebook-deplatforming-militia-violence>.

260. See *supra* note 71.

reports shape understandings of transparency and the way “success” and “failure” are understood and measured in content moderation. It suggests we take a thousand-foot view—a bean counter’s view—asking “how much,” while directing attention away from qualitative measurements centered on the substantive impact of content moderation practices. The statistics reported provide no insight into concerns about inconsistent or biased applications or outcomes of rules. In this way, transparency reports perform a sort of “transparency washing,”²⁶¹ directing analysis in a certain way to answer certain questions, while at the same time obscuring data necessary to answer the substantive normative questions that undermine a platforms’ ability to legitimately perform the application subfunction.

By design, transparency reports provide a limited and skewed glimpse into corporate practice. The reports provide a very partial picture of content removals and strategically divert attention away from the substance of corporate content moderation policies, the procedures for removals pursuant to them, and the substantive impact of those removals on speakers and listeners. This redirection of attention is consistent with platforms’ overall interest in limiting the extent to which the public and policymakers were aware of their vast and largely unchecked moderation activities, particularly those involving greater human review.²⁶²

Through the lens of the resolution or application subfunctions, then, transparency reports do not fill the gaps that undermine platforms’ ability to perform it legitimately. The statistics do not provide substantive information about the who and why of content moderation, nor do they help answer important questions about procedural regularity and consistent, impartial application of rules. Even with respect to the narrow substantive task of holding governments to account, researchers believe they are ineffective.²⁶³ Industry insiders have mixed views on their substantive impact. Some find that some governments were emboldened by data about the number and success of other governments’ requests for data and content removal. Others reported that that transparency about government requests improved government

261. Zalnieriute, *supra* note 257, at 139.

262. See FARLETON GILLESPIE, *CUSTODIANS OF THE INTERNET: PLATFORMS, CONTENT MODERATION, AND THE HIDDEN DECISIONS THAT SHAPE SOCIAL MEDIA passim* (2018) (describing social media platforms efforts to limit public awareness of content moderation through strategies including emphasizing the role of algorithms in content moderation, and keeping human workers in the shadows).

263. Losey, *supra* note 65, at 3456 (concluding data in transparency reports are insufficient for oversight or debate and urging companies to increase the granularity of reported data and provide greater specificity about legal processes governments use).

adherence to rule of law, evident in narrower requests with clear legal basis and greater procedural regularity.

Transparency reports could, if designed differently, address a systemic failing in content moderation systems and practices: the lack of visibility into what is removed, on what basis, and at whose request. In doing so, they could provide a means of holding companies accountable for maintaining standards of procedural regularity, consistent application of rules, and fair and equal treatment—thereby helping the public hold governments to account for speech removal practices. In their current design, however, they operate more as a symbol than substantive protection.

The strategic role that transparency reports play underscore their largely, though not purely, symbolic nature. At their worst, they do not establish the trustworthiness or legitimacy of firms to perform the moderating function²⁶⁴ but rather operate as transparency theatre, diverting attention to other actors and away from platforms. Yet, they are a symbolic legal structure with promise. Their evolution from information about government and third-party removals to information about removals under corporate rules is promising. However, moving from performative transparency to symbolic legal structures that align content moderation practices with democratic norms requires access to qualitative data and more detailed statistical data.

While each of these three structures—transparency reports, Google’s Advisory Council, and Facebook’s Oversight Board—respond to some extent to the values at stake in the relevant subfunction at issue, none robustly integrates the competencies necessary to render the exercise of platform power substantively and procedurally legitimate. Analyzing these structures in relation to content moderation subfunctions exposes their limitations and, as we develop in the next Section, offer a tool for reconfiguring content moderation to yield more legitimate distributed content governance frameworks.

B. THE CONSTRUCTIVE TURN: USING A FUNCTIONAL FRAMEWORK TO CONFIGURE CONTENT MODERATION

As its application to these examples demonstrates, a focus on subfunctions and constraints helps surface the ways that content moderation has been (and can be) organized and identifies the potential social and political implications of those design choices. In doing so, it clarifies why particular content moderation regimes draw particular, and particularly vociferous, objections. More importantly, consistent with the aims of the New Governance research

264. *Id.*

described above,²⁶⁵ the functional typology can assist regulators and other stakeholders in constructing content moderation regimes to align competencies with subfunctions and to couple them with constraints that further align delegations with public values. It provides a playbook, or a set of design patterns,²⁶⁶ to assist us in coupling allocations with constraints to build content moderation systems that adhere to public values, despite being largely composed of private entities. Below we walk through an example that illustrates the benefits of the framework in making this constructive turn.

1. *Leveraging Competencies and Addressing Democratic Deficits: Reimagining the Global Internet Forum to Counter Terrorism*

To meet their commitment under the EU Code of Conduct on Countering Illegal Hate Speech Online, Facebook, Google, Microsoft, and Twitter committed to, among other things, create the Global Internet Forum to Counter Terrorism (GIFCT). A key activity of the GIFCT is maintaining a shared hash database of terrorist content.²⁶⁷ While initially developed by the four companies, it is now used by thirteen companies.

On the surface, this shared hash database might appear to be a reasonable extension of the approach pioneered with the NCMEC CSAM hash database described above. It too provides a common resource to support the *locating* subfunction and by doing so, reduce the cost—financial and human—of multiple companies, and individuals within them, having to review and *identify* some of the most gruesome and disturbing content on the web. Further, by pooling the knowledge of identified terrorist content and automating the locating function across different platforms, the GIFCT database can speed up the removal of content valorizing violence. Coordinated multi-platform action is considered exceedingly important to stop the spread of violent content that can spread virally, build visibility and support for terrorist ideologies, from white supremacy to religious extremism, and fuel copycat actions.

265. See *supra* text accompanying notes 84–89.

266. For an alternative way to think about the construction of a subset of content moderation policies, see generally DAPHNE KELLER, BUILD YOUR OWN INTERMEDIARY LIABILITY LAW: A KIT FOR POLICY WONKS OF ALL AGES (2019).

267. In 2016, Facebook, Microsoft, Twitter, and YouTube announced they would work together to create a shared industry database of online terrorist content. *Partnering to Help Curb Spread of Online Terrorist Content*, FACEBOOK (Dec. 5, 2016), <https://about.fb.com/news/2016/12/partnering-to-help-curb-spread-of-online-terrorist-content>. This database later became part of the Global Internet Forum to Counter Terrorism. *About*, GLOB. INTERNET F. TO COUNTER TERRORISM, <https://gifct.org/about> (last visited Mar. 24, 2022).

Despite these similarities to the NCMEC hash database, the GIFTC database has been roundly criticized. The variety of concerns and objections are captured in a letter sent to the EU objecting to the then-draft EU proposal for regulating the dissemination of terrorist content online by an impressive and large group of the human rights organizations, journalists associations, and researchers.²⁶⁸ They noted that the definition of terrorist content is unstable, unshared, and subject to exceptions that protect important public accountability functions including news reporting and human rights documentation.²⁶⁹ Absent a shared and stable definition, using a set of distributive identified materials to drive removals across platforms and jurisdictions raised substantial concerns. In addition, given the importance of contextual evaluation, they raised concerns about the use of automated content moderation tools, in particular upload filters.²⁷⁰ Given the substantial risk of over removal, they objected to the profound lack of transparency and accuracy that generally attends automated decision-making.²⁷¹ They argued that “because it is impossible for automated tools to consistently differentiate activism, counter-speech, and satire about terrorism from content considered terrorism itself, increased automation will ultimately result in the removal of

268. Access Now et al., *Joint Letter to EU Parliament: Vote Against Proposed Terrorist Content Online Regulation* (Mar. 25, 2021), <https://www.hrw.org/news/2021/03/25/joint-letter-eu-parliament-vote-against-proposed-terrorist-content-online#>. The regulation, adopted in June 2021, underwent substantial amendments, several of which addressed concerns raised by the group. Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online, 2021 O.J. (L 172) 79. While it maintains the one-hour time period for hosting service providers to remove terrorist content upon notice by a competent authority, it allows for delay for “objectively justifiable technical or operational reasons,” and it requires annual reports by governments and platforms about removals, user notification of content removal determinations, and appeals processes, uses the definition of terrorist content from the existing Directive (EU) 2017/541 of the European Parliament and of the Council of 15 March 2017 on combating terrorism, 2017 O.J. (L 88) 6, and includes exceptions for content distributed for journalistic, research, or artistic purposes. See Katrien Luyten, *Addressing the dissemination of terrorist content online*, EUR. PARLIAMENTARY RSCH. SERV. (July 15, 2021), [https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI\(2020\)649326](https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2020)649326) (describing legislative process, including stakeholder engagements, and final proposal); *Terrorist Content Online*, (Apr. 2021), https://ec.europa.eu/home-affairs/document/download/506de61d-c53f-489f-a9e3-e16e78793e81_en. Relatedly, the French Constitutional Council, which reviews the constitutionality of legislation, struck down key provisions of the French Law on Countering Online Hatred, because they required platforms to make decisions about the legality of content without judicial involvement. Aurelien Breeden, *French Court Strikes Down Most of Online Hate Speech Law*, N.Y. TIMES (June 18, 2020), <https://www.nytimes.com/2020/06/18/world/europe/france-internet-hate-speech-regulation.html>.

269. Access Now et al., *supra* note 267.

270. *Id.*

271. *Id.*

legal content like news content and content about discriminatory treatment of minorities and underrepresented groups.”²⁷² They objected to the lack of procedural constraints, writing that “[t]he lack of judicial oversight is a severe risk to freedom of expression, assembly, association, religion and access to information.”²⁷³ Finally, given the lack of agreement on definitions, they raised concerns about a system that would paper over conflicting local laws and norms and allow “one Member State [to] extend its enforcement jurisdiction beyond its territory without prior judicial review and consideration for the rights of individuals in the affected jurisdictions.”²⁷⁴

These critiques elucidate how subject-matter-specific attributes of policy definition and application combine with the specific allocations of subfunctions in the GIFCT to undermine its substantive and procedural legitimacy.

The CSAM database is run and populated by NCMEC, a nonprofit entity that holds expertise in child sexual exploitation and is independent from the commercial platforms.²⁷⁵ NCMEC has particular operational capacity—it is funded and legally authorized to do things other entities cannot—and reflects the interests of victims and the common interest in eradicating both CSAM and the underlying activity it captures.²⁷⁶ NCMEC is tasked with applying the definition to content it receives to determine whether it should be added to the CSAM database. The hashes in that database, already determined to meet the definition of child sexual abuse material, are then used to identify and locate images that match (or nearly match) them on participating services. The definition of CSAM is established in public law in the United States, a definition that is consistent and at times identical to those in other countries. However, as described above, to avoid edge cases, NCMEC has chosen to limit the CSAM in the hash database to images considered the worst of the worst. CSAM material is illegal without exception. There is no journalistic, research, or other exceptions that make its storage, viewing, or distribution legal. For these reasons cultural and contextual cues are largely irrelevant to the application of the rule, limiting the potential gap between identification and application. To the extent that content can be classified on its four corners, it is deemed to meet the definition.

For these reasons, identification of this more limited set of CSAM should be both relatively straightforward and relatively consistent across platforms—

272. *Id.*

273. *Id.*

274. *Id.*

275. *See supra* text accompanying notes 134–139.

276. *Id.*

informed by shared law rather than by variable corporate policy. In this context, relying on a shared set of identified materials should pose little additional risk of over removal because identified content is likely to be subject to moderation under the rules. Finally, technology is used only to locate content that matches human-labeled content, not to independently identify new CSAM to which the definition has not yet been applied. As a system, this functional arrangement aligns subfunctions with competencies and builds in constraints (the narrower set of images included) to address potential risks and leverages technical actors (automation) in a discrete manner.

The configuration of the GIFCT Shared Industry Hash Database differs from the CSAM structure in fundamental ways. First, there is no shared public definition to guide the GIFCT's actions. There is no globally accepted agreement on the definition of terrorist content, and company standards vary. The lack of a shared and stable definition raises concerns about the appropriateness of pooling content identified across platforms and, importantly, across borders and targeting it for removal. In addition, unlike CSAM, there is a recognized need for exceptions. For example, terrorist content may be posted, shared, and stored to document human rights violations or to report on them. Applying policies to discrete pieces of content that may be embedded in larger works—archives of abuse or news stories, for example—makes judgments outside the context important to interpret the meaning of content in this area. Thus, with respect to terrorist content there is likely to be a substantial gap between content identified as potentially actionable and content found to be so after the rules are applied.

Second, the GIFCT is an industry consortium, not a non-profit with specific expertise, raising concerns about its independence, representativeness, and expertise. Secrecy around the structure of the work, including who identifies content as appropriate for inclusion in the database, and how GIFCT deals with the distinct policies of different companies, have furthered concerns about independence and expertise.

Finally, while technical actors are used to locate content that matches material previously identified for removal in ways similar to the CSAM database, the lack of clarity as to who determines what is in the database and what criteria are used to determine inclusion make this matching activity far more problematic. The technical means are the same, but the impact is decidedly different. In a partial effort to address this, initial corporate statements emphasized that matches against the database would not result in automatic removal; however, recent statements and reports indicate this is no

longer true: companies have backed away from their initial constraint of platform specific human review before removal.²⁷⁷

A functional approach suggests several ways subfunctions could be reconfigured to bring greater legitimacy into the content moderation regime for terrorist content. First, with regards to the task of defining, governments could establish a stable public agreement on the definition of terrorism content.²⁷⁸ The GIFCT, like NCMEC, moreover, could publicly agree to use the shared identifying and locating resource of the database for a narrower subset of terrorist content. Second, the GIFCT could create a representative body of independent experts to develop implementation guidelines for this narrower set of content. Third, the GIFCT could create an independent body to make determinations under this definition, limiting the ability of any platform to poison the system with poor evaluations and ensuring both expertise and independence in these decisions. This is of particular importance where liability regimes create risks to platforms—legal or other—that may incentivize over removal.

Finally, given the importance of context to the proper identification of terrorist content, the system should build in additional constraints, both technical and human. On the technical side, locating content could be constrained to ensure that material is not removed from known news sites or human rights archives, using things such as domain-level blocking. Unlike in the CSAM arena, where context, including the site on which content is located, is unimportant to whether it can be regulated, when dealing with terrorist content, context informs the legitimate application of the definition. The exception for content distributed for journalistic, research, or artistic purposes included in the EU regulation on the dissemination of terrorist content online underscores the potential importance of location and other factors that contribute to an understanding of purpose.²⁷⁹ In addition, the content of the database could be subject to review and audit by independent experts and coupled with broader transparency reports, like those envisioned by the Santa Clara Principles,²⁸⁰ to ensure automation does not come at the cost of speech

277. See BSR, 2021. “Human Rights Assessment: Global Internet Forum to Counter Terrorism” at 41.

278. See *id.* at 33 (concluding, after an assessment of GIFCT and consultation with stakeholders, that the task of creating a shared definition of terrorist and violent extremist content “properly resides with governments”).

279. Regulation (EU) 2021/784 art. 1, ¶ 3, 2021 O.J. (L 172) 79, 89.

280. *Santa Clara Principles 1.0*, THE SANTA CLARA PRINCIPLES ON TRANSPARENCY & ACCOUNTABILITY IN CONTENT MODERATION, <https://santaclaraprinciples.org/scp1/> (last visited Mar. 24, 2022). Business for Social Responsibility recommends that GIFCT figures out

that is particularly valuable, such as documenting human rights abuses and journalistic coverage of terrorist events.

Through reallocating and constraining subfunctions, the benefits of this shared identification and location resource could be realized in a manner that is substantively and procedurally legitimate.

V. CONCLUSION

In his forward-looking 1998 Article, *Lex Informatica: The Formulation of Information Policy Rules Through Technology*, to which this Issue is dedicated, Joel Reidenberg foresaw many of the elements that would come to characterize the regulatory environment of information infrastructures. The trans-jurisdictional scope of networks, he wrote, would decentralize the role of traditional government regulation in content governance, leading to fragmentation and confusion in rules about information flows. Sovereign laws could still wield some influence, but other forms of rulemaking and enforcement, including technical solutions and system design choices framed by private actors and those who designed technology for them, would shape policy decisions.

Our development of a functional framework for understanding, assessing, and constructing content moderation reflects Reidenberg's challenge to policymakers to "understand, consciously recognize, and encourage" the actual workings of these distributed forms of governance. Its identification of subfunctions supports a rigorous analysis of the way that content moderation actually functions on the ground, the concrete choices available regarding allocations of discrete subfunctions to different public or private actors, and how to leverage different actors' capacities and competencies. It surfaces the normative implications of different content moderation configurations and thereby facilitates an assessment of how such allocations can be constrained—either through processes or through limits on the use of automation—in ways that address those normative concerns.

Through this lens, a functional framework both offers a means for critically assessing the way various content moderation regimes allocate and constrain various subfunctions and generates constructive insights regarding the structuring of the content moderation function in new contexts. Its focus on the relevant content moderation subfunction involved in three structures that private platforms have adopted in response to legitimacy critiques—Google's Advisory Council on the Right to be Forgotten, Facebook's Oversight Board, and the use of transparency reports—surfaces the particular normative

how to enable annual, publicly reported, third-party reviews of the hash-sharing database. BSR, 2021. "Human Rights Assessment: Global Internet Forum to Counter Terrorism" at 44.

concerns implicated by the particular task and the governance competencies needed to address them. It thus enables a granular analysis of ways that the task is structured—to whom the subfunction is allocated, and how its exercise is constrained—and points to the ways that each fall short.

Looking forward, the functional framework permits a proactive analysis into how subfunctions in content moderation regimes might appropriately be structured, through law and private action, to promote legitimate governance systems in the future. It thus responds to Reidenberg's charge, providing a framework that regulators and others can use to construct distributed content governance regimes that restrain the power of platforms to pursue narrow self-interests while leveraging their capacity and expertise, along with that of other stakeholders, to advance the public interest.

AUTOMATED VIDEO INTERVIEWING AS THE NEW PHRENOLOGY

Ifeoma Ajunwa[†]

ABSTRACT

This Article deploys the new business practice of automated video interviewing as a case study to illuminate the limitations of traditional employment antidiscrimination laws. Employment antidiscrimination laws are inadequate to address the unlawful discrimination attributable to emerging workplace technologies which gatekeep employment opportunities. The Article maintains that the practice of automated video interviewing is based on shaky or unproven social scientific principles that disproportionately impact racial minorities. In this way, the practice of automated video interviewing is analogous to the pseudoscience of phrenology, which enabled societal and economic exclusion through the legitimization of eugenicist and racist attitudes. After parsing the limitations of traditional antidiscrimination law to curtail emerging workplace technologies such as video interviewing, this Article argues that ex ante legal regulations, such as those derived from the late Professor Joel Reidenberg's Lex Informatica framework, may be more effective than ex post remedies derived from the traditional employment antidiscrimination law regime.

The Article argues that one major benefit of applying a Lex Informatica framework to video interviewing is developing legislation that considers the technology's capabilities rather than how actors intend to use it. In the case of automated hiring, such an approach would mean actively using the Uniform Guideline on Employee Selection Procedures to govern the design of automated hiring systems. For example, the guidelines could dictate design features for the collection of personal information and treatment of content. Other frameworks, such as Professor Pamela Samuelson's "privacy as trade secrecy" approach could govern design features for how information from automated video interviewing systems may be transported and shared. Rather than reifying techno-solutionism, a focus on the technological capabilities of automated decision-making systems offers the opportunity for regulation to start at inception, which in turn could affect the design and development of the technology. This is a preemptive approach that sets standards for how the technology will be used and is a more proactive legal approach than merely addressing the negative consequences of the technology after they have occurred.

DOI: <https://doi.org/10.15779/Z38RX93F1Q>

© 2021 Ifeoma Ajunwa.

† Associate Professor of Law, University of North Carolina School of Law. Many thanks to the great late Professor Joel Reidenberg whose prescient scholarship informs this piece. Thanks also to my research assistants, Kylie McDonnell and Jake Schindler. A special thanks to the editors of the *Berkeley Technology Law Journal* for their fastidious edits.

TABLE OF CONTENTS

I.	INTRODUCTION	1175
II.	AUTOMATED VIDEO INTERVIEWING AS AI-ENABLED PHRENOLOGY	1178
A.	THE RISE OF AUTOMATED VIDEO INTERVIEWING	1178
B.	THE PHRENOLOGICAL ORIGINS OF AUTOMATED VIDEO INTERVIEWS	1182
1.	<i>The History of Phrenology</i>	1182
2.	<i>Phrenology in Law and Society</i>	1185
3.	<i>Phrenology and AI</i>	1187
C.	THE RACIAL IMPACT OF FACIAL AND EMOTION ANALYSIS	1190
D.	THE PSEUDOSCIENCE OF EMOTION RECOGNITION	1192
III.	LIMITATIONS IN LEGAL PROTECTIONS FOR APPLICANTS	1195
A.	TITLE VII	1195
B.	THE AMERICANS WITH DISABILITIES ACT (ADA)	1200
C.	PRIVACY LAW PROTECTION FOR JOB APPLICANTS?	1206
1.	<i>Notice and Consent</i>	1207
2.	<i>State Law</i>	1209
3.	<i>Fair Credit Report Act (FCRA) to the Rescue?</i>	1214
IV.	APPLYING A LEX INFORMATICA FRAMEWORK	1218
A.	TREATMENT OF CONTENT	1219
1.	<i>Criterion Validity for Automated Video Interviewing</i>	1220
2.	<i>Content Validity for Automated Video Interviewing</i>	1221
3.	<i>Construct Validity for Automated Video Interviewing</i>	1221
B.	TREATMENT OF PERSONAL INFORMATION	1222
C.	PRESERVATION OF OWNERSHIP RIGHTS: “PRIVACY AS TRADE SECURITY”	1222
V.	CONCLUSION	1224

I. INTRODUCTION

Jessica Clements, a job applicant with a visual impairment, had this to say about their automated video interview: “I couldn’t read the questions, I had to zoom in. And when it flipped to the front-facing camera, it was actually really distracting.”¹ Alex Huang, a job applicant and non-native English speaker, suspects he lost several job opportunities because automated video interviewing is prevalent in his job industry, financial services.² He believes that although he speaks fluent English, automated video interview systems had trouble understanding his tone and syntax.³ He finally got his current job position by insisting on an interview conducted by a human rather than an AI system.⁴ As AI-based video interviewing continues to grow as a prominent recruiting tool, it is critical to examine how such workplace technologies could serve as end runs against employment antidiscrimination laws such as Title VII of the Civil Rights Act of 1964⁵ and the Americans with Disabilities Act.⁶

This Article deploys automated video interviewing as a case study to consider the limitations of traditional employment antidiscrimination laws in addressing unlawful discrimination—particularly when such discrimination has been mediated by emerging workplace technologies such as automated hiring programs.⁷ In another law review article, I have noted how the adoption of automated technologies for hiring represents a paradox.⁸ The quandary is that automated technologies, which are often adopted as antibias interventions, have often been found to not only replicate the bias they were

1. Alex Lee, *An AI to Stop Hiring Bias Could Be Bad News for Disabled People*, WIRED (Nov. 26, 2019), <https://www.wired.co.uk/article/ai-hiring-bias-disabled-people>.

2. Telephone interview with Alex Huang (Aug. 6, 2020).

3. *Id.*

4. *Id.*

5. 42 U.S.C. §§ 2000e–2000e-17.

6. 42 U.S.C. §§ 12101–12213.

7. *See generally* James Grimmelman & Daniel Westreich, *Incomprehensible Discrimination*, 7 CAL. L. REV. ONLINE 164, 176 (2017) (discussing limited legal protections for job applicants confronted with automated hiring platforms); Pauline T. Kim, *Data-Driven Discrimination at Work*, 58 WM. & MARY L. REV. 857, 887–88 (2017); Charles A. Sullivan, *Employing AI*, 63 VILL. L. REV. 395, 395 (2018); Jeffrey M. Hirsch, *Future Work*, 2020 U. ILL. L. REV. 889, 939–41 (2020); Ifeoma Ajunwa, *An Auditing Imperative for Automated Hiring Systems*, 34 HARV. J.L. & TECH. 621 (2021).

8. *See generally* Ifeoma Ajunwa, *The Paradox of Automation as Anti-Bias Intervention*, 41 CARDOZO L. REV. 1671, 1679 (2021) (arguing that existing disparate impact legal frameworks are inadequate to address algorithm-based discrimination and advocating for a *discrimination per se* cause of action under Title VII, which would allow plaintiffs to “assert that a hiring practice . . . is so egregious as to . . . shift the burden of proof . . . to the defendant . . . to show that its practice is non-discriminatory.”).

meant to evade but, in fact, also amplify it.⁹ Confounding the issue is that the deployment of technology intermediaries renders bias both harder to discover and even more onerous to prove.¹⁰

This Article is descriptive as it illuminates the legal problems associated with automated video interviewing as a business practice. The Article is also prescriptive as it charts a way towards redress that expands our understanding of available legal remedies. To do this, the Article brings together two important fields of legal literature, integrating law and technology scholarship and employment law scholarship. Employment law scholarship has often been preoccupied with ex post adversarial measures for addressing unlawful discrimination.¹¹ Law and technology scholarship, on the other hand, has evolved to focus on ex ante collaborative methods such as auditing regimes and design features modification to address potential discrimination.¹² In the case of emerging workplace technologies, such as video interviewing, this Article operates from the normative viewpoint that ex ante legal regulations, such as those derived from the late Professor Joel Reidenberg's *Lex Informatica* framework,¹³ are more effective than ex post remedies derived from the traditional employment antidiscrimination law regime.

This Article proffers two important contributions. First, in line with my previous articles on emerging workplace technologies, it continues to challenge the received wisdom that AI technologies deployed in the workplace could be devoid of human bias.¹⁴ Not only have several real-world findings proven this assumption to be false,¹⁵ the nature and function of emerging AI technologies

9. *Id.*

10. *See generally* Ajunwa, *supra* note 7, at 639 (arguing that automated hiring technology furthers employment discrimination while masking such discrimination through opaque, unaccountable algorithms).

11. *See* Ajunwa, *supra* note 8, at 1685.

12. *See id.*; *see also* Kim, *supra* note 7, at 901; Solon Barocas & Andrew Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 718–19 (2016).

13. *See* Joel R. Reidenberg, *Lex Informatica: The Formulation of Information Policy Rules Through Technology*, 76 TEX. L. REV. 553, 581 (1998) (comparing some current legal policy areas to challenges faced by early merchants and arguing that the body of law created by those merchants, known as “Lex Mercatoria,” could be used to regulate information flows in the digital age); *see also infra* Part IV.

14. *See* Ajunwa, *supra* note 8, at 1671 (“A received wisdom is that automated decision-making serves as an anti-bias intervention. The conceit is that removing humans from the decision-making process will also eliminate human bias. The paradox, however, is that in some instances, automated decision-making has served to replicate and amplify bias.”); *see also* Ajunwa, *supra* note 7, at 679.

15. *See generally* Jeffrey Dastin, *Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women*, REUTERS (Oct. 10, 2018, 7:04 PM), <https://www.reuters.com/article/us->

intimate that the opposite is true; there is convincing evidence that emerging workplace technologies can amplify, routinize, and obscure unlawful employment discrimination.¹⁶ Given this known sociotechnical phenomenon, it behooves policymakers to attend to the potential for unlawful discrimination occasioned by the use of these new technologies.

A second contribution of this Article is an invitation to policymakers and employment antidiscrimination law advocates to join in dialogue with law and technology scholars. Too often, law is conceptualized as if it exists in a vacuum, and inadequate attention is paid to societal forces. The Article invites the readers to contemplate how law and technology are co-constitutive.¹⁷ Much like the law holds the power to constrain technological innovations, similarly, the capabilities of emerging technologies should inform new legal regimes. Using automated video interviewing as case study, this Article adds to the growing legal scholarship on novel legal frameworks to address new controversies of law wrought by new technical inventions.¹⁸

The roadmap for the Article is as follows: Part I charts the rise of automated video interviewing as a business practice rooted in the drive for efficiency in the hiring process. It also describes the phrenological origins of automated video interviewing. Part II tracks the limitations of legal protections

amazon-com-jobs-automation-insight/amazon-scrap-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G (detailing the discovery of an Amazon-built automated hiring system that turned out to be discriminatory against women); Sheridan Wall & Hilke Schellmann, *LinkedIn's Job Matching AI Was Biased. The Company's Solution? More AI*, MIT TECH. REV. (June 23, 2021), <https://www.technologyreview.com/2021/06/23/1026825/linkedin-ai-bias-ziprecruiter-monster-artificial-intelligence/>; Lydia Dishman, *The Bias You Didn't Know Existed in Job Ads and Recruiting Software*, FAST CO. (Sept. 7, 2015), <https://www.fastcompany.com/3051182/the-bias-you-didnt-know-existed-in-job-ads-and-recruiting-software>.

16. See Barocas & Selbst, *supra* note 12, at 720; Ajunwa, *supra* note 7, at 630 (noting that automated hiring platforms, in particular, hold the potential to both amplify and obfuscate bias).

17. See generally Lauren Edelman, *When Organizations Rule: Judicial Deference to Institutionalized Employment Structures*, 117 AM. J. SOCIO. 888 (2011) (noting the ways that organizational development is influenced by the law and how organizations can in turn influence law); JULIE E. COHEN, *BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTION OF INFORMATIONAL CAPITALISM* (2019) (arguing that informational capitalism has developed amid loopholes in extant law and also now seeks to influence future law).

18. Some choice examples include both literature in employment and labor law, including Kim, *supra* note 7 and Hirsch, *supra* note 7, and also, beyond: Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218 (2019); Margot E. Kaminski, *The Right to An Explanation, Explained*, 34 BERKELEY TECH. L.J. 189 (2019); Sonia K. Katyal, *Private Accountability in The Age of Artificial Intelligence*, 66 UCLA L. REV. 54 (2019); Devin R. Desai & Joshua A. Kroll, *Trust But Verify: A Guide to Algorithms & The Law*, 31 HARV. J.L. & TECH. 1 (2018).

for applicants subjected to automated video interviewing by examining extant antidiscrimination laws found in Title VII of the Civil Rights Act, the Americans with Disabilities Act (ADA), privacy law, and the Federal Credit Reporting Act (FCRA). Part III examines what applying a Lex Informatica framework to automated video interviewing might entail. It focuses on the treatment of content, personal information, and the preservation of ownership rights for data collected as part of the automated interviewing process.

II. AUTOMATED VIDEO INTERVIEWING AS AI-ENABLED PHRENOLOGY

This Part discusses the rise of video interviewing as a business practice. It also discusses the phrenological origins of automated video interviewing, the limitations of AI for facial analysis and emotion detection, and what this means for racial exclusion in the workplace.

A. THE RISE OF AUTOMATED VIDEO INTERVIEWING

Video interviewing without the use of artificial intelligence emerged as a field in the early 2000s with two companies: HireVue and Montage.¹⁹ According to VidCruiter CEO Sean Fahey, these companies “invented pre-recorded video interviewing, where people record at home while hiring managers were doing something else.”²⁰ In its nascency, convenience was the driving force behind video interviewing. Employers would provide candidates with a set of standardized questions; candidates would then video record their answers to each question; then employers would review those answers to make hiring decisions. This system meant employers could better compare candidates and get multiple opinions while saving money and time on recruiter travel.²¹ Around the same time, other interview-focused hiring startups were in the works, such as GreenJobInterview which provided one of the first purpose-built, live video interview platforms before the age of Zoom.²²

Since its inception, video interviewing has evolved to include the use of artificial intelligence. Asynchronous interviews recorded and sent via the internet have now become standard practice as opposed to mailing physical tapes. Starting in 2015, HireVue became one of the first to offer AI-based

19. *Where Did Video Hiring Come From and Where Is It Going?*, VIDCRUITER, <https://vidcruiter.com/video-interviewing/history-of-video-interview/#:~:text=In%20HireVue's%20case%2C%20founder%20Mark,managers%20were%20doing%20something%20else.%E2%80%9D> (last visited Nov. 6, 2021).

20. *Id.*

21. HireVue, *The HireVue Story—Mark Newman*, VIMEO (Apr. 26, 2013, 6:03 PM), <https://vimeo.com/64921188>.

22. VIDCRUITER, *supra* note 19.

assessments.²³ HireVue's assessments traditionally used vocal and facial analysis technology, drawing on "[a] database of about 25,000 pieces of facial and linguistic information," to provide recruiters with a measure of a candidate's potential job performance.²⁴ As of 2019, the algorithms assessed factors such as "a candidate's tone of voice, their use of passive or active words, sentence length and the speed they talk," and facial expressions such as "brow furrowing, brow raising, the amount eyes widen or close, lip tightening, chin raising and smiling."²⁵ Although HireVue claims it discontinued use of its facial recognition technology as of 2021,²⁶ it continues to use lingual analysis and there is no formal ban or rule preventing the reimplementation of facial recognition technology for automated hiring systems.²⁷ HireVue's marriage of AI and video interviewing has become entrenched as a standard practice for recruitment. A 2020 study that analyzed the claims and practices of various algorithmic hiring companies found that one-third of the eighteen companies analyzed deployed video-based assessments.²⁸

Video interview adoption on the whole is rising alongside the integration of AI assessments. In 2011, a survey of 506 companies found that 47% were using video interviewing to speed up the hiring process, while another 22% responded they would consider video interviewing as a tool to recruit geographically diverse candidates.²⁹ A 2015 survey of 700 executives found

23. See generally *Our Science: Meet the IO Psychology Team*, HIREVUE, <https://www.hirevue.com/our-science> (last visited Apr. 22, 2021) (noting the use of IO psychology as part of AI-based interview assessments).

24. Ivan Manokha, *How Using Facial Analysis in Job Interviews Could Reinforce Inequality*, PBS NEWS HOUR (Oct. 7, 2019, 3:26 PM), <https://www.pbs.org/newshour/economy/making-sense/how-using-facial-recognition-in-job-interviews-could-reinforce-inequality#:~:text=The%20technology%2C%20developed%20by%20US,are%20videoed%20answering%20identical%20questions.&text=HireVue%20says%20it%20speeds%20up,the%20speed%20of%20information%20processing>.

25. *Id.*

26. See Roy Maurer, *HireVue Discontinues Facial Analysis Screening*, SOC'Y FOR HUM. RES. MGMT. (Feb. 3, 2021), <https://www.shrm.org/resourcesandtools/hr-topics/talent-acquisition/pages/hirevue-discontinues-facial-analysis-screening.aspx>.

27. Cf. Will Knight, *Job Screening Service Halts Facial Analysis of Applicants*, WIRED (Jan. 12, 2021), <https://www.wired.com/story/job-screening-service-halts-facial-analysis-applicants/> (noting some state statutes and local level ordinances that ban facial recognition).

28. Manish Raghavan, Solon Barocas, Jon Kleinberg & Karen Levy, *Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices*, 2020 CONF. ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY (FAT*) 469, 473 (2020).

29. Heather O'Neill, *Video Interviewing Cuts Costs, but Bias Worries Linger*, WORKFORCE (Oct. 5, 2011), <https://www.workforce.com/news/video-interviewing-cuts-costs-but-bias-worries-linger>.

that 50% were using video interviews to “narrow the candidate pool.”³⁰ Significantly, these statistics report the state of video interviewing *before* AI-based assessments were introduced; recent reports suggest the industry has grown at rapid rates since the integration of this new AI-based assessment technology. As of 2018, 60% of organizations were using video interviews, a number which dramatically spiked in 2020 due to global shutdowns prompted by the COVID-19 pandemic.³¹ A 2020 Gartner HR survey reported that 86% of respondents were turning to new virtual interview technology to facilitate remote hiring.³² Data specific to industry leader HireVue shows that 733 corporations were using the platform as of 2021, the majority of which employed more than 10,000 employees and touted more than \$1 billion a year in revenue.³³ Computer software, health care, retail, and financial services industries ranked among the top users of HireVue’s services.³⁴ The expansive reach of HireVue’s automated video hiring technology as a single platform warrants closer scrutiny. A serious consideration of AI-based video interview technology’s potentially discriminatory effects is necessary given the rapid adoption and scale of video interviewing. This need for scrutiny is heightened by the unproven validity of its AI-based predictions.

Such an examination is especially salient as anecdotal evidence suggests that the corporate frenzy to join the bandwagon of technological hiring comes at the expense of candidates who have the most to lose, such as racial minorities and applicants with disabilities. Even the average job candidate has decried the lack of autonomy and the depersonalization when the hiring decision hinges on AI-based video interview technology. A group of candidates interviewed by the Washington Post shared that they found the video interview process “alienating and dehumanizing.”³⁵ Some candidates

30. See Roy Maurer, *Use of Video Interviewing Continues to Grow*, SOC’Y FOR HUM. RES. MGMT. (Aug. 21, 2015), <https://www.shrm.org/resourcesandtools/hr-topics/talent-acquisition/pages/use-video-recruiting-grow.aspx>.

31. Nilam Oswal, *The Latest Recruitment Technology Trends and How to Really Use Them*, PC WORLD (Feb. 9, 2018, 4:56 PM), <https://www.pcworld.idg.com.au/article/633219/latest-recruitment-technology-trends-how-really-use-them/>; *Gartner HR Survey Shows 86% of Organizations Are Conducting Virtual Interviews to Hire Candidates During Coronavirus Pandemic*, GARTNER: NEWSROOM (Apr. 30, 2020), <https://www.gartner.com/en/newsroom/press-releases/2020-04-30-gartner-hr-survey-shows-86-of-organizations-are-cond>.

32. GARTNER: NEWSROOM, *supra* note 31.

33. *Companies Using HireVue*, ENLYFT, <https://enlyft.com/tech/products/hirevue#:~:text=We%20have%20data%20on%20733,and%20%3E1000M%20dollars%20in%20revenue> (last visited Apr. 22, 2021).

34. *Id.*

35. Drew Harwell, *A Face-Scanning Algorithm Increasingly Decides Whether You Deserve The Job*, WASH. POST (Nov. 6, 2019, 12:21 PM), <https://www.washingtonpost.com/technology/>

have even begun to distinguish between video interviews and “real” interviews, which they believe are more likely to lead to actual hiring decisions. A college graduate speaking to Slate magazine described feeling disappointed when she realized her interview was a taping to be reviewed by “an A.I. thing” as opposed to a conversation with a real recruiter.³⁶ For some candidates, video interviews have come to represent part of the culling process of automated hiring rather than a meaningful opportunity to prove one’s qualifications.³⁷

For other candidates, the concern around video interview technology runs even deeper. Kat, a software engineering student also interviewed by Slate magazine, noted that video interviewing technology made her “[feel] like [she] was not valued as a human.”³⁸ Also, as a Black woman, her concerns around dehumanization were exacerbated by her recognition that “A.I. is known to perpetuate bias against people of color or fail to recognize them at all.”³⁹ Her friends and other professionals encouraged her to decline video interviews which used AI, which she reported was her plan going forward.⁴⁰ Individuals with disabilities have received the same advice from disability advocates and companies alike who believe video interview algorithms run the risk of screening out candidates with disabilities.⁴¹

Although HireVue and some other video interview platforms claim to mitigate bias in their hiring systems, concerns for race, disability, and other types of discrimination do not seem unfounded. Not all companies engage in algorithm de-biasing efforts, and for those that do, it is not clear that those mitigation efforts are effective for catching all manifestations of bias in hiring algorithms.⁴² Given the presence of instances like Amazon’s proprietary screening tool that systematically discriminated against women,⁴³ there is ample evidence of AI-based hiring technologies perpetuating discrimination in the past, and it is thus critical to identify the shortcomings of existing

2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job/.

36. Rachel Withers, *Should Robots Be Conducting Job Interviews?*, SLATE (Oct. 5, 2020, 9:00 AM), <https://slate.com/technology/2020/10/artificial-intelligence-job-interviews.html>.

37. Ajunwa, *supra* note 7, at 622–23.

38. Withers, *supra* note 36.

39. *Id.*

40. *Id.*

41. Jim Fruchterman & Joan Mellea, *Expanding Employment Success for People with Disabilities*, BENETECH 1, 3 (Nov. 2018), <https://benetech.org/wp-content/uploads/2018/11/Tech-and-Disability-Employment-Report-November-2018.pdf>.

42. See Raghavan et al., *supra* note 28, at 477–78 (noting that simply using outcome based de-biasing to comply with the Title VII 4/5 rule may not effectively capture all forms of algorithmic discrimination).

43. See Dastin, *supra* note 15.

legislation in order to design a more robust, protective regulatory regime for automated video interviewing.

B. THE PHRENOLOGICAL ORIGINS OF AUTOMATED VIDEO INTERVIEWS

The problems with automated video interviewing are manifold. To fully comprehend the contours of the problem, a ground level examination of the origins of automated video interviewing is necessary. The premise that emotions or character may be surmised from the human face or facial expressions is part of the theoretical scaffolding for automated video interviewing. Scholars like Kate Crawford have characterized the thinking behind this premise as a “phrenological impulse”—the desire to entertain assumptions about an individual’s emotions and character merely from external appearances.⁴⁴ Yet, there remains no scientific consensus that artificial intelligence systems are capable of accurately interpreting human emotions from facial expressions. In fact, one psychological study, conducted in 2019, found no strong evidence that artificial intelligence could accurately ascertain a person’s emotions solely from facial expressions.⁴⁵ Yet, despite the lack of evidence to support the efficacy of automated video interviewing, this type of recruitment technology is rapidly becoming entrenched as part of hiring and recruitment efforts.⁴⁶

1. *The History of Phrenology*

Starting roughly in the 1800s, phrenology was the brainchild of German physiologist Dr. Franz Joseph Gall.⁴⁷ Gall, a relatively “handsome man” who himself had a “broad ‘noble head,’” sought to transform the then somewhat informal field of psychology, the study of the human mind and its functions, into a true science.⁴⁸ Gall rejected the traditional philosophical grounds for understanding the human brain, instead turning to what he considered to be concrete scientific data in order to create his own scientific hypothesis for

44. Kate Crawford, *Time To Regulate AI That Interprets Human Emotions*, NATURE (Apr. 6, 2021), https://www.nature.com/articles/d41586-021-00868-5?utm_source=twitter&utm_medium=social&utm_content=organic&utm_campaign=NGMT_USG_JC01_GL_Nature.

45. Lisa Feldman Barrett, Ralph Adolphs, Stacy Marsella, Aleix M. Martinez & Seth D. Pollak, *Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements*, 20 PSYCH. SCI. PUB. INT. 1, 46–47 (2019).

46. See, e.g., *HireVue Ranked A Fastest Growing Company on Deloitte’s 2018 Technology Fast 500*, HIREVUE (Nov. 15, 2018), <https://www.hirevue.com/press-release/deloitte-2018-technology-fast-500-hirevue-ranked-fastest-growing-company> (demonstrating the reach of AI interviewing technology).

47. Pierre Schlag, *Commentary: Law & Phrenology*, 110 HARV. L. REV. 877, 878 (1997).

48. *Id.* (quoting NELSON SIZER, FORTY YEARS IN PHRENOLOGY 380–81 (1888)).

inner workings of the mind—a field that would come to be known as phrenology.⁴⁹ Gall began his studies by observing animal and human behavior. He studied social structures, from “family life” to “jails and asylums,” intending to identify mankind’s “fundamental faculties” through an analysis of objective data.⁵⁰ At the center of Gall’s theory was the cerebral localization hypothesis.⁵¹ Gall believed that the brain was divided into separate essential organs, each of which served a unique and essential function.⁵² These functions extended not only to essential intellectual skills but also genetically coded for moral and emotional capabilities.⁵³ Thus, according to Gall’s hypothesis, one’s individual behavior and intellect was directly related to the development and structure of one’s physical brain anatomy. Gall in part rested his argument on an appeal to the specialization seen throughout nature: just as eyes serve a specialized function, just as ears serve a predetermined purpose, it followed logically, for Gall, that different regions of the brain would follow suit.⁵⁴ Gall based many of his “objective” scientific observations on anecdotal evidence. For example, Gall decided that those who “learn by heart” always feature “large prominent eyes.”⁵⁵ Gall concluded that there was a similar pattern with other physical traits corresponding to mental capacities. According to Gall, these observations soon led him to believe with certainty “that the difference in the form of heads is occasioned by the difference in the form of the brains.”⁵⁶ Also, these differences presented themselves by size. Any region of the brain that was more developed in an individual would be larger in size and would also feature more prominently in outward appearance.⁵⁷

Based upon his cerebral localization hypothesis, Gall conducted further pseudoscientific tests to arrive at a construction of no less than “twenty-seven fundamental faculties” located in different regions of the skull that explained human behavior.⁵⁸ Gall relied on “empirical observation” of numerous

49. *Id.*

50. *Id.* at 879.

51. *Id.*

52. *Id.*

53. *Id.*

54. *Id.*

55. *Id.* at 880.

56. *Id.*

57. *Id.*

58. *Id.* (enumerating Gall’s list of 27 faculties as follows: Amativeness, Philoprogenitiveness, Adhesiveness, Combativeness, Destructiveness, Secretiveness, Acquisitiveness, Self-Esteem, Love of Approbation, Cautiousness, Eventuality [and Individuality], Locality, Form, Vocabulary, Language, Coloring, Tune, Number, Constructiveness, Comparison, Causality, Wit, Ideality, Benevolence, Imitation, Veneration, & Firmness).

individuals, depending heavily on correlation to associate some specific traits with particular regions of the brain.⁵⁹ After identifying where they deemed particular faculties to reside in the brain, Gall and other phrenologists made value judgments concerning what each faculty meant for individual behaviors and personalities. Phrenologists took great care to isolate various faculties from others, describing the detailed nature of each region of the brain and the nuanced observations and implications used to determine each region's independent relationship to behavior.⁶⁰ Building on Gall's work, Dr. Johann Spurzheim observed that the faculties Gall identified could be further divided into feelings and intellect, with subdivisions that created a detailed hierarchy and organizational framework within which to interpret phrenological findings.⁶¹

Although the study of phrenology began in the late eighteenth century, it did not gain prominence until 1815 when a review in the *Edinburgh Review*, a respected intellectual magazine of the time, condemned the newfound science as “utterly destitute of every qualification necessary for the conduct of a philosophical investigation.”⁶² Middle-class individuals, fascinated by this new and previously unknown theory, however, began to follow phrenologists' findings despite backlash from the scientific community.⁶³ In the eyes of the public at that time, Spurzheim had successfully refuted the *Edinburgh Review's* critiques, and in 1820, the first phrenological society was formed in Edinburgh, Scotland.⁶⁴ Phrenology's popularity became a wave that swept from Britain to America. In 1838, the first meeting of the Phrenological Association convened; it was modeled on respected scientific associations which had excluded phrenology from its ranks.⁶⁵ In phrenology, the “enthusiastic and the arrogant” found a “scientific” justification for personally held beliefs, from Christians to radicals and racists.⁶⁶ The American “phrenological Fowlers” were a group of phrenology advocates who gave lectures, established institutions across the world, published new research, and even read heads for a fee.⁶⁷ By 1844, the Fowlers' publishing house was distributing phrenological research and

59. *Id.*

60. *Id.* at 882.

61. *Id.* at 883.

62. John van Wyhe, *Ridiculing Phrenology: “This Persecuted Science”*, THE HIST. PHRENOLOGY ON THE WEB, <http://www.historyofphrenology.org.uk/ridicule.htm> (last visited Jan. 4, 2022).

63. John van Wyhe, *Overview*, THE HIST. PHRENOLOGY ON THE WEB, <http://www.historyofphrenology.org.uk/overview.htm> (last visited Jan. 4, 2022).

64. *Id.*

65. *Id.*

66. *Id.*

67. *Id.*

propaganda nationwide.⁶⁸ Fowler philosophy, spearheaded by Lorenzo Niles Fowler and Samuel R. Wells, based their version of phrenology on the theories of George Combe.⁶⁹ Though many original phrenologists viewed the Fowlers' brand of phrenology as a deviated form, it was this brand that embedded itself in American social and legal institutions and which persisted despite scientific invalidation.⁷⁰

2. *Phrenology in Law and Society*

Phrenology in America embedded itself in both fashionable and legal society, no doubt thanks to the Fowlers' work. In 1873, writer Mark Twain underwent a secret phrenological examination, wherein a phrenologist determined he had no sense of humor; interestingly, a few months later when Twain returned and publicly announced his well-known name, the phrenologist discovered Twain's skull did in fact house an impressive bump of humor.⁷¹ Although inconsistencies such as those identified during Twain's encounter were well known during the height of phrenology's popularity, public opinion far outweighed scientific criticisms. Workplace evaluations and screening decisions deployed phrenology, with George Combe himself stating he "would never employ a clerk who had not a large coronal region."⁷² Similar to many personality tests of the early twentieth century, employers occasionally used phrenology to determine an individual's suitability for a given career;⁷³ indeed, some employers would include particular phrenological profiles in their job solicitations.⁷⁴ For example, in 1912, a phrenological evaluation of a seven-year-old suggested she was best suited for a career in medicine or teaching.⁷⁵

Phrenology also made its way into the American legal system: it informed theories of criminal law reform, it was a method by which jurists evaluated an individual's culpability, and it was critically used as a "mitigating factor" during criminal sentencing.⁷⁶ Phrenological evaluations were used to determine who

68. *Id.*

69. *Id.*

70. *Id.*

71. Amanda C. Pustilnik, *Violence on The Brain: A Critique of Neuroscience in Criminal Law*, 44 WAKE FOREST L. REV. 183, 191 (2009).

72. Olivia Goldhill, *Centuries Before Myers-Briggs, Workplace Personalities Were Assessed Using Skull Measurements*, QUARTZ AT WORK (Dec. 29, 2017), <https://qz.com/work/1168283/centuries-before-myers-briggs-workplace-personalities-were-assessed-using-phrenology/>.

73. *Id.*

74. Minna Scherlinder Morse, *Facing a Bumpy History*, SMITHSONIAN MAG. (Oct. 1997), <https://www.smithsonianmag.com/history/facing-a-bumpy-history-144497373/>.

75. Goldhill, *supra* note 72.

76. Pustilnik, *supra* note 71, at 192–93.

may be at risk of committing a crime; some police departments used phrenology to typify criminals and arrest them, even in the absence of any evidence a crime had been committed.⁷⁷ Even judges relied on phrenological evidence as fact in official judgments. In the 1853 murder trial *Farrer v. State*, the Ohio Supreme Court relied on phrenological evaluation to determine whether a housekeeper could be held liable for poisoning a youth. The judge presiding over the case ruled that the housekeeper was “remarkably ugly,” thus it was evident that she was both “criminally insane” and subject to “murderous impulses.”⁷⁸ In 1840, a judge stated from the bench that “‘no man . . . would dispute that the brain . . . consists of distinct organs, each having a distinct function, and that power of function is influenced by organic size.’”⁷⁹ Phrenology even influenced the M’Naghten test for insanity, which assumes that an individual’s “ability to know right from wrong” is distinct from any mental disease they may suffer.⁸⁰ This distinction was erroneously rooted in phrenology’s theory of separate, individually functioning mental organs; yet it nevertheless persisted in case law all the way until 1966—well after phrenology had fallen out of use.⁸¹

By the 1950s, the field of phrenology was all but dead. Its demise began nearly as soon as its success. In 1838, at the same time the first phrenological society was called to order, scientists already had evidence that the brain did not actually house enough separate regions to allow each major personality trait its own organ.⁸² In fact, evidence at that time increasingly suggested that many parts of the brain must work in tandem to function.⁸³ Eventually, scientists also came to realize that brain size had little to no correlation with intelligence or efficiency.⁸⁴ Thus, for much of phrenology’s rise there was growing evidence of its invalidity.

Despite these challenges, the practice of phrenology persevered for over a century, embedded in social institutions and common thought. By 1888, the theory was so ingrained that the editors of *Encyclopedia Britannica* felt the need to publish a seven-page essay to refute the theory.⁸⁵ Thus, it was not so much scientific evidence alone that finally led society to cast phrenology aside as

77. *Id.* at 192.

78. *Id.* at 193–194.

79. *Id.* at 194.

80. *Id.* at 193.

81. *Id.*

82. *Id.* at 194.

83. *Id.*

84. *Id.*

85. Morse, *supra* note 74.

phrenology had never truly even been classified as a science.⁸⁶ It was instead a combination of changing social theories and norms paired with scientific evidence that resulted in genuine change.⁸⁷ Simply put, phrenology became “unfashionable.”⁸⁸ As Freudian psychoanalysis gained popularity in the early 1900s, people began to abandon the theory of fixed traits in favor of the more intriguing and mysterious influence of the unconscious mind.⁸⁹ Although the field of phrenology itself was eventually associated with “zealous extremists,” some of its influences lived on.⁹⁰ In the early twentieth century, the spirit of phrenology gave rise to the racially charged anthropological theory that Europeans were superior to other humans based on the shape and size of their skulls.⁹¹ Advocates for this spin-off movement included Paul Broca, who went on to found the Anthropological Society in Paris circa 1859.⁹²

3. *Phrenology and AI*

As the pseudoscience of phrenology fell into disuse, its influence did not merely cease. Rather, its core ideologies evolved into new, more fashionable theories. Phrenology was based on the fundamental belief that human behavior is innate—that an individual is born with certain set behavioral tendencies and capabilities.⁹³ Thus, one conclusion is that, at its core, phrenology presupposes human behavior as quantifiable. Phrenology presupposes that a limited number of behavioral traits exist and that the prevalence of such traits in an individual was directly proportional to their physical characteristics.⁹⁴ Franz Gall had sought to create a system wherein individuals could be objectively measured through quantifiable observations that would allow for useful systematic comparison. Gall’s methods were of course flawed.⁹⁵ He based his science on judgments derived from normative comparisons and perceived value.⁹⁶ Although Gall’s motivation was to create an objective study of the human mind, the result was a theory that was at best pseudoscientific and, at worst, a subjective social tool used to reinforce a static social hierarchy and rationalize class inequality.⁹⁷ These ideologies,

86. van Wyhe, *supra* note 63.

87. Pustilnik, *supra* note 71, at 194.

88. van Wyhe, *supra* note 63.

89. Pustilnik, *supra* note 71, at 194.

90. van Wyhe, *supra* note 63.

91. *Id.*

92. *Id.*

93. Schlag, *supra* note 47, at 879.

94. *Id.*

95. *Id.*

96. *Id.*

97. Goldhill, *supra* note 72.

motivations, and social implications did not disappear when phrenology fell out of use.

The ideological goal to objectively understand human behavior morphed into the well-respected field of psychology, which has experienced its own evolutions over the past century. Phrenology (1840s) was subsumed into the science of behaviorism (1920s), which sought to understand human motivation through observable behavior as opposed to observable physical features.⁹⁸ Cognitive psychology replaced behaviorism (1950s), which shifted focus from observable external behavior to observable brain functions in order to understand human traits such as perception, memory, problem-solving, and intelligence.⁹⁹ While the scientific evidence has shifted and methods of scientific evidence have improved, there remain some parallels between phrenological theory and cognitive psychology's focus on the brain's inner structures.¹⁰⁰

At the turn of the twenty-first century, a new product burst onto an already booming technology scene with the promise to improve efficiency in the hiring process: the digital interview.¹⁰¹ The digital interview initially presented a simple concept: allow individuals to access interview questions at home, prerecord their response, and save companies valuable time and resources.¹⁰² However, digital interview techniques have evolved into the present day automated video interviewing systems that do not merely passively record a candidate's response.¹⁰³ Rather, the systems are often the intermediate arbiters of the candidate's character and job suitability.¹⁰⁴ Today, the most popular and widely used of these technologies is HireVue, ranked one of the 500 fastest growing technology companies in 2018.¹⁰⁵ Starting in 2013, HireVue began using AI to enhance the video interview process.¹⁰⁶ HireVue's systems measure candidates' body language, tone, key words, and even, previously, facial expressions; the results come in the form of a single "employability score,"

98. Kendra Cherry, *The Origins of Psychology*, VERYWELLMIND (June 25, 2020), <https://www.verywellmind.com/a-brief-history-of-psychology-through-the-years-2795245>.

99. *Id.*

100. *Id.*

101. VIDCRUITER, *supra* note 19.

102. *Id.*

103. Richard Feloni, *I Tried the Software that Uses AI to Scan Applicants for Companies Like Goldman Sachs and Unilever Before Meeting Them—and It's Not as Creepy as It Sounds*, BUS. INSIDER (Aug. 23, 2017, 12:00 PM), <https://www.businessinsider.com/hirevue-ai-powered-job-interview-platform-2017-8> (containing information that demonstrates that the hiring software does evaluate applicants).

104. *Id.*

105. HIREVUE, *supra* note 46.

106. Feloni, *supra* note 103.

which is then ranked against other applicants.¹⁰⁷ HireVue claims its technology removes bias from the hiring process by applying a single, objective algorithm to all candidates, allowing for a fair evaluation.¹⁰⁸ Although in January of 2021, after criticism, HireVue announced that it will halt the use of facial analysis, it still retains the use of intonation and body language to make hiring decisions.¹⁰⁹ Thus, some might consider HireVue's system an iteration of phrenology.

Yet, in the same way that phrenology “scientifically” assessed individuals based on normative, anecdotal observations, video interview technology like HireVue's measures candidates' responses against a normative sample of individuals who are perceived to be successful at a given job. Companies such as HireVue collect “training data” in the form of interviews and performance records from existing high-performing employees at a given company.¹¹⁰ As training data, the traits exhibited by top performers are set as the variables that the automated systems will use to screen candidates. These traits are often nuanced—for example, better enunciation or simply leaning forward on the table could be traits that correlate to successful salespeople at a given company, rendering these as variables that the algorithm rewards.¹¹¹ This type of training data reflects one of the most basic logical fallacies: correlation is not causation.

Furthermore, basing an employment decision on correlations may undermine the equal opportunity principle. For example, Amazon took a candidate screening technology it had developed out of service once it came to light that the technology disproportionately ranked men higher than women.¹¹² One reason for this was that the technology, which operated on correlations, likely compared candidates to the traits commonly shared by top performers, and the top performers were overwhelmingly men. This is not because men were more competent but simply the result of past biases in recruitment that gave men a historical advantage. In this case, the averred objective scientific measurement was in actuality algorithmic processes reflecting societal biases—as the pseudoscience of phrenology had previously done.

107. Angela Chen, *The AI Hiring Industry Is Under Scrutiny—But It'll Be Hard to Fix*, MIT TECH. REV. (Nov. 7, 2019), <https://www.technologyreview.com/2019/11/07/75194/hirevue-ai-automated-hiring-discrimination-ftc-epic-bias/>.

108. Feloni, *supra* note 103.

109. Knight, *supra* note 27.

110. Alex Engler, *For Some Employment Algorithms, Disability Discrimination by Default*, BROOKINGS INST. (Oct. 31, 2019), <https://www.brookings.edu/blog/techtank/2019/10/31/for-some-employment-algorithms-disability-discrimination-by-default/>.

111. *Id.*

112. Dastin, *supra* note 15.

Akin to how phrenology sought to quantify human character through observable, physical traits, video interview technologies also seek to quantify and objectively understand human behavior as it relates to job success. An inherent underlying assumption of these technologies is that there exist observable, physical manifestations that give insight into the character and behavioral traits that define a successful individual. Video interviewing technology is purportedly motivated by objectivity,¹¹³ yet it ranks candidates based on judgments rooted in normative comparisons. The automated video interviewing algorithms are trained to search for certain traits deemed to be valuable, but these normative conclusions are based on samples of existing employees, and these samples are not random and may not be representative.¹¹⁴ Yet, society has uncritically embraced video interviewing technology much in the same way that it embraced phrenology. As of 2017, HireVue alone boasted more than 600 clients, many of them multinational corporations such as Unilever, Goldman Sachs, and Under Armor.¹¹⁵ Unfortunately, just like phrenological thinkers of the eighteenth century, early adopters of automated video interviewing have largely failed to consider the scientifically shaky foundations undergirding video interviewing technologies.

C. THE RACIAL IMPACT OF FACIAL AND EMOTION ANALYSIS

Although companies like HireVue claim to no longer use facial recognition,¹¹⁶ there have been no independent audits to substantiate such claims.¹¹⁷ Thus, it remains urgent to understand and redress the racial impact of both the facial and emotion analysis used for automated video interviews. In *Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices*, Raghavan and his coauthors point to “[a] wave of studies [which have] shown that several commercially available facial analysis techniques suffer from disparities in error rates across gender and racial lines.”¹¹⁸ In 2018, Joy Buolamwini and Timnit

113. Feloni, *supra* note 103.

114. *See id.*

115. *Id.*

116. *Hirevue Leads the Industry with Commitment to Transparent and Ethical Use of AI in Hiring*, HIREVUE (Jan. 12, 2021), <https://www.hirevue.com/press-release/hirevue-leads-the-industry-with-commitment-to-transparent-and-ethical-use-of-ai-in-hiring> (“Independently, early in 2020, HireVue proactively removed the visual analysis component from all of its new assessments. HireVue’s internal research demonstrated that recent advances in natural language processing had significantly increased the predictive power of language. With these advances, visual analysis no longer significantly added value to assessments.”).

117. Ajunwa, *supra* note 7, at 672 (noting that HireVue’s audit was conducted by a company that HireVue had hired, and after the audit was completed, there were still “many questions [left] unanswered”).

118. Raghavan, *supra* note 28, at 475.

Gebru examined the performance of facial analysis algorithms across four “intersectional subgroups” of males or females featuring lighter or darker skin.¹¹⁹ Buolamwini and Gebru found that algorithms designed to identify gender performed better on male faces as opposed to female and performed better on light faces as opposed to dark. Darker females were also the most misclassified of all groups.¹²⁰ This troubling finding suggests that the facial analysis software that video interview algorithms employ may be less accurate when identifying job candidates of color and women.

This finding is further corroborated by Lauren Rhue, who found that the emotion analysis feature of two facial recognition algorithms “interprets emotions differently based on the person’s race.”¹²¹ One recognition software interpreted Black individuals as angrier than White individuals regardless of whether the individual was smiling; the other platform viewed Black individuals as more contemptuous than White individuals when their face featured an “ambiguous” expression, though “[a]s the players’ smile widens, the disparity disappears.”¹²² Thus, not only is facial recognition technology less accurate at identifying women and people with darker skin tones, it is also less accurate at interpreting emotions that individuals with dark skin express. Given video interviewing’s reliance on facial and emotion recognition technology, this disparity is troubling. In 2019, HireVue claimed their algorithm assessed such nuanced traits as “brow furrowing, brow raising, the amount eyes widen or close, lip tightening, chin raising and smiling.”¹²³ However, AI is clearly unable to accurately identify and assess the meanings of these subtle facial movements for all groups. Thus, by building algorithms to rely on this inaccurate technology, video interview platforms are nearly guaranteeing discriminatory results.

For AI scholar Luke Stark, these discriminatory outcomes are not merely a byproduct of flawed design; rather, the racialization of the human face is integral to the mission of facial analysis.¹²⁴ By “attach[ing] numerical values to the human face,” humans are necessarily being quantified and judged by classifiable visual signs—race foremost of all.¹²⁵ Thus, for Stark, facial

119. Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROC. MACH. LEARNING RSCH. 1, 2 (2018).

120. *Id.* at 8.

121. Lauren Rhue, *Racial Influence on Automated Perceptions of Emotions*, RACE, AI, & EMOTIONS 1, 1 (2018).

122. *Id.*

123. Manokha, *supra* note 24.

124. See Luke Stark, *Facial Recognition Is the Plutonium of AI*, 25 XRDS: CROSSROADS 50, 53 (2019).

125. *Id.* at 52.

recognition technologies “both create and reinforce discredited categorizations around gender and race.”¹²⁶ It is this observation that leads Stark to boldly claim “facial recognition is the plutonium of AI . . . anathema to the health of human society, and [something that should be] heavily restricted as result.”¹²⁷ Legal scholars like Woodrow Hartzog have also called for a wide ban on facial recognition technologies.¹²⁸

Interviews are the gateway to work and earning a livelihood, a fundamental human right.¹²⁹ Incorporating facial and emotion recognition technology into the interview process means the interview process could become tainted by racialized bias. To date, millions of video interviews relying on facial and emotion analysis have been conducted.¹³⁰ Although some platforms have claimed to discontinue the use of facial recognition technology,¹³¹ many automated systems still claim to act as emotion recognition systems.

D. THE PSEUDOSCIENCE OF EMOTION RECOGNITION

As succinctly put in *Emotional Entanglement: China’s Emotion Recognition Market and Its Implications for Human Rights*—a recent report on the background, uses, and ethical issues that underlie the use of emotion recognition technologies in China’s authoritarian state—“[t]wo fundamental assumptions undergird emotion recognition technologies: that it is possible to gauge a person’s inner emotions from their external expressions, and that such inner emotions are both discrete [that is quantifiable] and uniformly expressed across the world.”¹³² These core principles have historical precedent. In an examination of modern emotion-recognition technology, Rich Firth-Godbehere asserts that the roots of universal emotion theories trace all the way back to the Greek philosopher Aristotle and seventeenth century artist Charles Le Brun, a proponent of the controversial and since discredited racist

126. *Id.*

127. *Id.*

128. See Evan Selinger & Woodrow Hartzog, *The Inconsentability of Facial Surveillance*, 66 LOY. L. REV. 101, 105 (2019); see also CLIC Faculty, *Professor Woodrow Hartzog Calls for a Ban on Facial Recognition Technology in New Publication*, NORTHEASTERN: CLIC (Apr. 14, 2020), <https://www.northeastern.edu/clic/2020/04/>.

129. See G.A. Res. 217 (III) A, Universal Declaration of Human Rights, art. 23(1) (Dec. 10, 1948) (“Everyone has the right to work, to free choice of employment, to just and favourable conditions of work and to protection against unemployment.”).

130. See Manokha, *supra* note 24 (discussing HireVue’s 2019 claim that their algorithm assessed such nuanced traits as “brow furrowing, brow raising, the amount eyes widen or close, lip tightening, chin raising and smiling.”); see also Maurer, *supra* note 26.

131. See Maurer, *supra* note 26.

132. ARTICLE 19, ENTANGLEMENT: CHINA’S EMOTION RECOGNITION MARKET AND ITS IMPLICATIONS FOR HUMAN RIGHTS 15 (2021), <https://www.article19.org/wp-content/uploads/2021/01/ER-Tech-China-Report.pdf>.

field of physiognomy.¹³³ In the late nineteenth century, Charles Darwin sought to marry theories of universal emotions with the science of his day. Darwin published *The Expression of the Emotions in Man and Animals* in 1872, an offshoot of evolutionary theory that suggested “some kind of common evolutionary ancestor” explained the parallels between “some instinctual actions” both animals and humans express.¹³⁴ However, it was not Darwin but mid-twentieth century psychologist Paul Ekman whose research would lay the groundwork for emotion recognition AI.

In the 1960s, Ekman laid the groundwork for “Basic Emotion Theory (BET)” around two primary theories.¹³⁵ Firstly, Ekman, working in conjunction with scientist Silvan Tomkins and Wallace Friesen, theorized that there are “six basic emotions: happiness, anger, sadness, disgust, surprise, and fear.”¹³⁶ Conducting studies with a remote civilization in Papua New Guinea, Ekman determined these emotions are consistent “across cultures.”¹³⁷ Ekman’s second theory concerned “micro expressions.” He believed that not only are basic emotions universal but also that minute expressions, which “occur briefly in response to stimuli, are signs of ‘involuntary emotional leakage [which] exposes a person’s true emotions.’”¹³⁸

According to AI scholar Kate Crawford, this is the pseudoscience underlying today’s emotion recognition AI.¹³⁹ Crawford argues that the marriage of Ekman’s theories with computer science was one of convenience: “the six emotions Ekman described fit perfectly into the model of the emerging field of computer vision.”¹⁴⁰ Ekman’s theory was attractive because it allowed emotions to be quantified and, in turn, allowed the quantification of emotions to be “standardized and automated at scale.”¹⁴¹ However, this union wholly ignored the mounting questions regarding the validity and accuracy of Ekman’s theories.

133. Rich Firth-Godbehere, *Silicon Valley Thinks Everyone Feels the Same Six Emotions*, NEXT (Sept. 5, 2018), <https://qz.com/1392130/silicon-valley-thinks-everyone-feels-the-same-six-emotions/>

134. *Id.*

135. ARTICLE 19, *supra* note 132, at 15.

136. Firth-Godbehere, *supra* note 133.

137. ARTICLE 19, *supra* note 132, at 15.

138. *Id.*

139. Crawford, *supra* note 44.

140. *Id.*

141. *Id.*

Since the twentieth century, scientists had already begun to suspect that theories of universal emotion expression were inaccurate.¹⁴² Anthropologist Margaret Mead was an early skeptic of this idea. Researching the people of a remote Samoan island in the 1920s, Mead concluded “that fundamental human experiences—including emotions—varied from culture to culture.”¹⁴³ Ekman’s New Guinea studies appeared to challenge Mead’s conclusions. Yet, his research methods were later called into question. For one, the fact that it was later discovered that the subjects of Ekman’s study had in actuality previously interacted with Western researchers before, called into question the extent to which they were truly culturally isolated.¹⁴⁴ Moreover, Ekman’s use of translators and photographs of exaggerated faces also called into question the accuracy of his findings; more recent research shows that emotions are harder to recognize when less exaggerated.¹⁴⁵ Given Ekman’s flawed methodology, it is not clear his findings actually challenge Mead’s conclusions concerning cultural differences in human emotional expression.

Ekman’s theory of micro expressions has also been proven “to be both unreliable (due to [the] brevity and infrequency [of micro expressions]) and discriminatory.”¹⁴⁶ This finding is particularly discrediting for video interviewing technology, which is known to rely on facial analysis that analyzes minute facial movements.¹⁴⁷ Another study found that facial expressions and the universal emotions that supposedly underlie them are “only weakly associated” at best.¹⁴⁸ Furthermore, in perhaps one of the more compelling recent studies, *Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements*, Lisa Feldman Barrett and her coauthors, “systematic[ally] review” evidence concerning emotion recognition to ultimately conclude that “how people communicate anger, disgust, fear,

142. See generally Lisa Feldman Barrett, *Are Emotions Natural Kinds?*, 1 PERSPS. ON PSYCHOL. SCI. 28, (2006) (discussing several studies disproving the universality of facial expressions as representing emotions).

143. Firth-Godbehere, *supra* note 133.

144. *Id.*

145. *Id.* (citing Copernicus Center for Interdisciplinary Studies, *Language, Emotion, and Facial Expression—J. Russell*, YOUTUBE (Nov. 26, 2011), <https://www.youtube.com/watch?v=oS1ZtvrgDLM>).

146. ARTICLE 19, *supra* note 132, at 16 (citing Andrea Korte, *Facial Recognition Technology Cannot Read Emotions, Scientists Say*, AM. ASS’N FOR THE ADVANCEMENT OF SCI. (Feb. 17, 2020), <https://www.aaas.org/news/facial-recognition-technology-cannot-read-emotions-scientists-say>).

147. See Manokha, *supra* note 24 (discussing video interview platform HireVue’s facial analysis technology, which formerly considered facial movements such as “brow furrowing, brow raising, the amount eyes widen or close, lip tightening, chin raising and smiling”).

148. ARTICLE 19, *supra* note 132, at 16 (citing J.A. RUSSELL & J.M. FERNÁNDEZ-DOLS, COHERENCE BETWEEN EMOTIONS AND FACIAL EXPRESSIONS (2017)).

happiness, sadness, and surprise varies substantially across cultures, situations, and even across people within a single situation.”¹⁴⁹ As Barrett and her coauthors note, emotion recognition literature is lacking context-specific studies of facial expressions.¹⁵⁰ They propose that the unknowns around facial recognition technology should cause scientists to “step back from what we think we know about reading emotions in faces.”¹⁵¹ This warning should also be heeded by automated video interview system creators. Essentially, Barrett and her coauthors’ work invalidates automated video interviewing systems, not merely on the idea of biased algorithms or biased training data but on the premise that the entire field of science on which automated video interviewing is based is misleading. According to the authors:

[T]ech companies may well be asking a question that is fundamentally wrong. Efforts to simply ‘read out’ people’s internal states from an analysis of their facial movements alone, without considering various aspects of context, are at best incomplete and at worst entirely lack validity, no matter how sophisticated the computational algorithms.¹⁵²

III. LIMITATIONS IN LEGAL PROTECTIONS FOR APPLICANTS

Given the deeply flawed science behind automated video interviewing and the growing evidence that its use may perpetuate racial and other biases in hiring, one question remains: what legal protections, if any, are afforded to the job candidate who encounters automated video interviewing as a part of their job search? In the Sections below, the Article examines the extant legal protections available to job applicants and parse their limitations apropos automated video interviewing.

A. TITLE VII

Title VII of the Civil Rights Act of 1964 “prohibits employers from discriminating against employees and applicants for employment on the bases of race, color, religion, national origin, and sex.”¹⁵³ Video interviewing algorithms may run afoul of Title VII if algorithmic decision-making is found to discriminate against candidates across any of these protected classes. For

149. Barrett et al., *supra* note 45, at 1.

150. *Id.* at 48.

151. *Id.* at 51.

152. *Id.* at 48.

153. Jennifer Issacs, *Proving Title VII Discrimination in 2019*, AM. BAR ASS’N, https://www.americanbar.org/groups/young_lawyers/projects/no-limits/proving-title-vii-discrimination-in-2019/ (last visited Mar. 12, 2021).

example, algorithms relying on biased or incomplete training data may produce discriminatory hiring decisions that penalize those who do not reflect the White male majority that has historically held an advantage in the workplace.

A Title VII claim brought against a discriminatory video interview algorithm would likely follow the path of a disparate impact claim as opposed to disparate treatment. Intent is essential to disparate treatment claims, and it would be particularly difficult to prove intent when the machine acts as an opaque intermediary between employers and candidates.¹⁵⁴ Intent may be especially difficult to prove in cases where video interview tools take steps to screen out bias on the basis of protected classes. Even if an employee could prove intent and harm under disparate treatment theory, an employer may still then claim a legitimate, nondiscriminatory alternative reason for its action.¹⁵⁵ Plausible alternatives include a significant correlation between the tool in question and job performance.¹⁵⁶

Unfortunately, disparate impact theory offers only a slightly better protection. In McKenzie Raub's Title VII analysis of video interview algorithms in *Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability*, she suggests that plaintiffs may have issues establishing a prima facie case under disparate impact theory "when the discrimination is the result of incomplete, incorrect, or non-representative data . . . [or data that] fails to represent groups in accurate proportions."¹⁵⁷ According to Raub, statistically proving discrimination, as required for a prima facie case, could be particularly complicated considering "segments of protected classes could be excluded from employment opportunities because of a lack of access to the required technology to participate in the hiring practices that use artificial intelligence."¹⁵⁸ Applying this insight specifically to video-based hiring, a minority applicant opt-out bias may mean that individuals who try to bring an adverse impact claim do not have enough peers who have used the technology to effectively prove their discrimination was statistical

154. See Stephanie Bornstein, *Antidiscriminatory Algorithms*, 70 ALA. L. REV. 519, 524 (2018) ("Worse still, current scholarship suggests, the apparent neutrality of algorithms and the 'black box' nature of machine learning make this hiring trend a new way of doing business that could be unreachable by existing antidiscrimination law."); McKenzie Raub, *Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices*, 71 ARK. L. REV. 529, 550 n.174 (2018).

155. See Jennifer Jolly-Ryan, *Have a Job to Get a Job: Disparate Treatment and Disparate Impact of the 'Currently Employed' Requirement*, 18 MICH. J. RACE & L. 189, 201 (2012).

156. See Sullivan, *supra* note 7, at 420–21; Sandra F. Sperino, *Disparate Impact of Negative Impact: Future of Non-Intentional Discrimination Claims Brought by the Elderly*, 13 ELDER L.J. 339, 358 (2005).

157. Raub, *supra* note 154, at 547–48.

158. *Id.*

rather than circumstantial. Thus, although a discriminatory video interview algorithm may in fact have an adverse impact, a lack of aggregated evidence may make it difficult for employees to establish a case for protection under Title VII.

Notably, other scholars take a slightly different approach from Raub on this issue by placing the onus on employers to prevent algorithmic discrimination.¹⁵⁹ As those scholars argue, “employment antidiscrimination law imposes an affirmative duty of care on employers to ensure that they are avoiding practices that would constrain equal opportunity in employment.”¹⁶⁰ Calling on the work of other legal scholars, I have argued that this duty, emanating from Title VII protections, would entail an “auditing imperative” for video interviewing.¹⁶¹ Such an imperative would require employers to proactively audit their algorithms for any instance of bias, which would in turn “enable litigation by generating data to serve as statistical evidence of disparate impact or by discovering practices that could be considered *discrimination per se*.”¹⁶² An auditing imperative could therefore aid plaintiffs in bringing an effective prima facie case under Title VII.

However, even if a job applicant can prove a prima facie case, it may be relatively easy for employers to establish that their criteria for the algorithmic models in question are job related and constitute a business necessity. Establishing a business necessity reason for the hiring practice can serve as an affirmative defense for employment discrimination.¹⁶³ As previously noted, for issues concerning artificial intelligence, the primary question “seems to be ‘whether . . . the target variable . . . is job related’ . . . [and] actually predictive of the job related trait.”¹⁶⁴ Video interview algorithms “are prognostic by nature,” created for the sole purpose of identifying job-related traits.¹⁶⁵ Relying on biased input data—where a target variable is perhaps positively correlated with both successful job performance as well as historical discrimination—means that an employer may meet its burden to prove that a model correlates to job performance even if the model has a discriminatory impact. The open

159. Ajunwa, *supra* note 7, at 626–27 (referencing the work of other legal scholars like Richard Thompson Ford, James Grimmelman, Robert Post, David Benjamin Oppenheimer, and Noah Zatz).

160. *Id.*

161. *Id.* at 625.

162. *Id.* at 674.

163. 42 U.S.C. § 2000e-2(k)(1)(A)(i) (“[A plaintiff] demonstrates that a respondent uses a particular employment practice that causes a disparate impact on the basis of [a protected characteristic] and the respondent fails to demonstrate that the challenged practice is job related for the position in question and consistent with its business necessity.”).

164. Raub, *supra* note 154, at 549.

165. *Id.* at 549–550.

question concerning an employer's burden of proof asks whether the model extends beyond proving mere statistical correlation to job performance; that is, as others have asked, does an employer have to go as far as to "[show] that no problems exist with the data or model construction that are biasing the results"?¹⁶⁶ The legal scholar Pauline Kim, suggests Title VII could be interpreted to apply this higher burden.¹⁶⁷ However, many other scholars suggest that it is unlikely the Court would require such proof under existing case law.¹⁶⁸ A plaintiff would need open access to an algorithm to parse out these insights themselves—and such access would almost certainly be impossible to obtain.¹⁶⁹ As Professor Sandra Sperino notes, employers are "reluctant to produce this information voluntarily," resulting in an informational asymmetry that disadvantages plaintiffs in the litigation process.¹⁷⁰

This power imbalance continues to play out even for plaintiffs who succeed to the next step in the litigatory process, when the claimant has the opportunity to prove that a less discriminatory alternative employment practice exists after an employer has made its case. As Raub points out, "[i]f an employer fails to effectively disclose or defend the validity of its algorithm and data collection . . . the plaintiff is hamstrung."¹⁷¹ That is, a claimant cannot effectively defend themselves against a model they cannot examine or understand. James Grimmelman and Daniel Westreich come to a similar conclusion in *Incomprehensible Discrimination*, wherein they examine the legal implications of a hiring model that is positively correlated to job performance yet yields a discriminatory impact.¹⁷² Grimmelman and Westreich find that it may be hard for a claimant to "improve on an algorithm it did not create and does not understand"; thus, the claimant would likely fail to offer the sort of "concrete and less discriminatory alternative" necessary to prevail under current Title VII case law.¹⁷³ Grimmelman and Westreich propose, like Pauline T. Kim, a heightened standard to prove business necessity, which would

166. *Id.* at 551 (quoting Kim, *supra* note 7, at 921).

167. Kim, *supra* note 7, at 921.

168. See Grimmelman & Westreich, *supra* note 7, at 168–69 (stating that an employer would theoretically meet its burden of proof "by showing an 'undisputed statistically and practically significant correlation' " between an algorithm's outcome and a measure of job performance); see also Barocas & Selbst, *supra* note 12, at 702–05 (highlighting that courts employ a varying standard of job-relatedness and business necessity and that courts generally accept some finding that an outcome is predictive of job-performance as satisfying an employer's burden).

169. Raub, *supra* note 154, at 550.

170. Sperino, *supra* note 156, at 361.

171. Raub, *supra* note 154, at 552.

172. Grimmelman & Westreich, *supra* note 7, at 164.

173. *Id.* at 169.

“[require an employer] to show not just that its model’s scores are . . . *correlated* with job performance but *explain* it.”¹⁷⁴ While such a standard may help a plaintiff prevail, this heightened standard is far from the standard interpretation of an employer’s burden under Title VII.

Take, for example, cases of accent discrimination. Accent discrimination is a credible threat of automated video systems given that many video interview algorithms employ vocal analysis. In fact, a recent audit of HireVue’s algorithms suggest that accent discrimination may already be present in the company’s assessment outcomes.¹⁷⁵ Title VII case law suggests that there is a path for candidates to bring such accent discrimination claims under Title VII’s “national origin” protection clause.¹⁷⁶ Under Title VII, an employer may only consider an employee’s accent when making a hiring decision “if [the] accent materially interferes with being able to do the job.”¹⁷⁷ Case law suggests that the mere presence of an accent alone does not rise to the level of material interference. In *Fragante v. Honolulu*, the Ninth Circuit distinguishes between discriminating against someone because an accent is present and discriminating on the grounds that an accent makes communication difficult.¹⁷⁸ An employer may only make a hiring decision based on the “effect” of a candidate’s accent.¹⁷⁹ Further, a manager’s subjective dislike or preference concerning an accent is likely not enough to prove material interference. In *EEOC v. Brown and Brown Chevrolet, Inc.*, the Equal Employment Opportunity Commission (EEOC) charged that a car dealership’s failure to promote a salesman on the grounds he should “speak ‘more like an American’” was a Title VII violation.¹⁸⁰ Algorithms which discriminate on the basis of accent would need to prove that the accent in question is a relevant factor in

174. *Id.* at 170.

175. Jeremy Kahn, *HireVue Drops Facial Monitoring Amid A.I. Algorithm Audit*, FORTUNE (Jan. 19, 2021, 12:01 PM), <https://fortune.com/2021/01/19/hirevue-drops-facial-monitoring-amid-a-i-algorithm-audit/>.

176. Mari J. Matsuda, *Voices of America: Accent, Antidiscrimination Law, and a Jurisprudence for the Last Reconstruction*, 100 YALE L.J. 1329, 1332 (1991).

177. *Fact Sheet: Immigrants’ Employment Rights Under Federal Anti-Discrimination Laws*, EEOC (Apr. 27, 2010), [https://www.eeoc.gov/laws/guidance/fact-sheet-immigrants-employment-rights-under-federal-anti-discrimination-laws#:~:text=Treating%20employees%20differently%20because%20they,able%20to%20do%20the%20job.&text=If%20a%20person%20has%20an,sh%20cannot%20be%20discriminated%20against](https://www.eeoc.gov/laws/guidance/fact-sheet-immigrants-employment-rights-under-federal-anti-discrimination-laws#:~:text=Treating%20employees%20differently%20because%20they,able%20to%20do%20the%20job.&text=If%20a%20person%20has%20an,sh%20cannot%20be%20discriminated%20against.).

178. *Fragante v. City & Cty. of Honolulu*, 888 F.2d 591, 599 (9th Cir. 1989).

179. *Id.*

180. David Woodfill, *Brown & Brown Settles Suit Over Nigerian Accent*, E. VALLEY TRIB. (Oct. 7, 2011), https://www.eastvalleytribune.com/news/brown-brown-settles-suit-over-nigerian-accent/article_f41851cd-f3ab-537b-9d60-74127f44a6ba.html (citing *EEOC v. Brown & Brown Chevrolet, Inc.*, No. CV-05-1575-PHX-ROS (filed D. Ariz. May 26, 2005)).

determining job performance. It is not clear that mere correlation between previous high performers is enough to meet this burden.

Yet, despite the potential for candidates to find protection under Title VII for instances of accent discrimination, the likelihood of prevailing remains low because employers are likely to mount a business necessity defense. As Mari Matsuda identifies in *Voices of America: Accent, Antidiscrimination Law, and a Jurisprudence for the Last Reconstruction*, in practice, “[t]he fact that communication is an important element of job performance . . . tends to trump this prohibition against discrimination, such that it is impossible to explain when or why plaintiffs will ever win in accent cases. In fact, they almost never do.”¹⁸¹ According to Matsuda, the issue is that Title VII prohibits discrimination on the basis of a protected class but allows discrimination on the basis of “job ability.”¹⁸² For accent discrimination, this means that when employers argue that accent is inextricably linked to job-related communication skills, they can effectively evade Title VII liability.¹⁸³ Matsuda summarizes the issue succinctly: “in every accent case the employer will raise the “[customers] ‘can’t understand [the employee or job candidate]’ defense, and in almost every reported case, the courts have accepted it.”¹⁸⁴ For video interview algorithms which show evidence of accent discrimination, this means that employers may effectively evade liability by claiming that the discrimination in question was a valid byproduct of the algorithm’s assessment of communication skills. Claiming the algorithm found that the applicant’s accent impeded effective communication with the AI in question may be enough for employers to prevail.

Overcoming the employer’s business necessity defense against a Title VII suit is incredibly difficult. Indeed, on the whole, Title VII places too great of a burden on plaintiffs to offer any substantive protection in the age of machine learning and video interviewing. Although there are mounting calls from some scholars to reconsider the mandates and burdens of Title VII in ways more favorable to plaintiffs, the current judicial interpretation of Title VII ultimately renders it inadequate to fully address the unlawfully discriminatory impact of video interviewing.

B. THE AMERICANS WITH DISABILITIES ACT (ADA)

Although the American with Disabilities Act (ADA) could provide some protection for disabled applicants, the heightened burden of proof for ADA

181. Matsuda, *supra* note 176, at 1332.

182. *Id.* at 1348.

183. *See id.* at 1350.

184. *Id.*

cases now established by *Murray*¹⁸⁵ means that proving discrimination on the basis of a disability for job applicants may be difficult. In 1990, Congress passed the ADA, a piece of civil rights legislation designed to explicitly encode the rights of disabled individuals in law.¹⁸⁶ Amended in 2008 to alter and significantly expand the definition of disability under the Act, the ADA applies to employers with fifteen or more employees and features specific protections for disabled individuals in various settings, including the job application process.¹⁸⁷ The ADA specifically regulates preemployment assessments, prohibiting the use of “qualification standards, employment tests or other selection criteria that screen out or tend to screen out an individual with a disability or a class of individuals with disabilities” unless the assessment or criterion is proven to be a job-related, business necessity.¹⁸⁸ As video interview algorithms serve as a form of assessment, they may therefore implicate the ADA if they are found to screen out applicants on the basis of their ability status. Employers must take care that their assessment algorithms allow employees with impairments concerning “sensory, manual, or speaking skills . . . [to achieve] results [that] accurately reflect the skills, aptitude, or whatever other factor of such applicant or employee that such test purports to measure.”¹⁸⁹ Failure to do so constitutes discrimination under the ADA.¹⁹⁰

Beyond preemployment assessments, the ADA also includes specific provisions concerning medical examinations. Although an employer is permitted to “make preemployment inquiries into the ability of an applicant to perform job-related functions,” the Act prohibits any medical examination or inquiry to determine an applicant’s disability status—be it in kind or severity—unless it constitutes a job-related, business necessity.¹⁹¹ Regardless of job-relatedness, the Act prohibits an employer from requiring any medical

185. *See Murray v. Mayo Clinic*, 934 F.3d 1101, 1105 (9th Cir. 2019), *cert. denied*, 140 S. Ct. 2720 (2020) (“Because *Head*’s reasoning is clearly irreconcilable with *Gross* and *Nassar*, we overrule *Head*’s holding that a plaintiff bringing a discrimination claim under Title I of the ADA need show only that a disability was a motivating factor of the adverse employment action. We hold instead that an ADA discrimination plaintiff bringing a claim under 42 U.S.C. § 12112 must show that the adverse employment action would not have occurred but for the disability.”).

186. *What is the Americans with Disabilities Act (ADA)?*, ADA NAT’L NETWORK, <https://adata.org/learn-about-ada> (last visited Mar. 17, 2021).

187. *Notice Concerning The Americans With Disabilities Act (ADA) Amendments Act of 2008*, EEOC (Mar. 25, 2011), <https://www.eeoc.gov/statutes/notice-concerning-americans-disabilities-act-ada-amendments-act-2008>; *Fact Sheet: Disability Discrimination*, EEOC (Jan. 15, 1997), <https://www.eeoc.gov/laws/guidance/fact-sheet-disability-discrimination>.

188. Americans with Disabilities Act of 1990, 42 U.S.C. § 12112(b)(6) (1990).

189. 42 U.S.C. § 12112(b)(7).

190. *Id.*

191. 42 U.S.C. §§ 12112(d)(2)(B), (d)(4)(A).

examinations before a conditional offer of employment is made.¹⁹² According to Melson-Silimon and her coauthors, in *Personality Testing and the Americans With Disabilities Act*, criteria for determining if a preemployment assessment constitutes a medical examination includes the following:

the test (a) was administered by a healthcare/medical professional; (b) was interpreted by a healthcare or medical professional; (c) was originally designed to reveal an impairment or an applicant's current mental or physical health; (d) was invasive; (e) measured a physiological response (e.g., heart rate) to a (job-related) physical task; (f) is typically used in a medical setting; or (g) involved the use of medical equipment.¹⁹³

Significantly, ADA provisions concerning medical examinations extend to “psychological tests that are designed to identify a mental disorder or impairment.”¹⁹⁴ The EEOC effectively distinguishes between prohibited psychological tests that constitute a medical examination and other forms of psychological tests; it states there are some permissible tests for pre-offer employment screening under the ADA, “includ[ing] measures of honesty, preferences, and habits.”¹⁹⁵

The EEOC acted in numerous cases since the ADA went into effect. When and how it chooses to enforce the ADA may offer significant guidance for interpreting how it may approach enforcement in terms of video interview algorithms. Take, for example, *EEOC v. Subway Inc.*, filed in Indiana.¹⁹⁶ The agency argued that the franchise “violated federal law by rejecting a hard-of-hearing applicant because of his hearing and resultant speech impairments.”¹⁹⁷ The franchise allegedly chose not to hire an impaired candidate “because of

192. *Questions and Answers: Enforcement Guidance on Disability Related Inquiries and Medical Examinations Under the Americans with Disabilities Act*, EEOC (July 27, 2000), <https://www.eeoc.gov/laws/guidance/questions-and-answers-enforcement-guidance-disability-related-inquiries-and-medical>.

193. Arturia Melson-Silimon, Alexandra M. Harris, Elizabeth L. Shoenfelt, Joshua D. Miller & Nathan T. Carter, *Personality Testing and the Americans With Disabilities Act: Cause for Concern As Normal and Abnormal Personality Models Are Integrated*, 12 INDUS. ORG. PSYCH. 119, 121 (2019) (citing *Enforcement Guidance on Disability-Related Inquiries and Medical Examinations of Employees Under the ADA*, EEOC (July 26, 2000), https://www.eeoc.gov/laws/guidance/enforcement-guidance-disability-related-inquiries-and-medical-examinations-employees#N_33_).

194. *Id.*

195. *Id.*

196. See Press Release, EEOC, Subway Franchisee to Pay \$28,700 to Settle EEOC Disability Discrimination Suit (Mar. 26, 2021), <https://www.eeoc.gov/newsroom/subway-franchisee-pay-28700-settle-eeoc-disability-discrimination-suit>.

197. *Subway Franchisee Sued by EEOC for Disability Discrimination*, EEOC (Sept. 23, 2020), <https://www.eeoc.gov/newsroom/subway-franchisee-sued-eeoc-disability-discrimination>.

his disability, citing a ‘communication concern’ due to the applicant’s ‘hearing’ and ‘speaking.’”¹⁹⁸ The EEOC argued that this adverse employment action constituted disability discrimination that violated the ADA.¹⁹⁹ This enforcement action is significant as it shows that the EEOC does not simply allow employers to argue that an impairment materially disqualifies a disabled individual from a given job. Although sandwich makers with specific ways of speaking may have been typical in the Subway franchise or may have even been preferable in the employer’s view, the EEOC effectively stated that such speaking patterns are not a legitimate consideration for job qualification such that the hard of hearing individual may be disqualified. Thus, video interview algorithms which consider certain speaking patterns in making an employment decision may directly violate the ADA.

In *EEOC v. Randstad, US, LP*²⁰⁰ the EEOC filed suit against a Maryland company that failed to hire an individual once he disclosed his autism.²⁰¹ The company allegedly initially considered the applicant highly qualified for the lab technician job in question, “fast-track[ing] [the candidate’s] participation in the hiring process” as result.²⁰² Once the applicant disclosed his disability, however, “he was told that the lab technician position had been put ‘on hold.’”²⁰³ Ultimately, the applicant was not hired and the company went on to fill the position with another recruit. The EEOC argued that this adverse employment decision was made in response to the applicant’s autism disclosure in violation of the ADA.²⁰⁴ The case was settled with Randstad agreeing to pay \$60,000.²⁰⁵ In the context of video interview assessments, this case is significant because it suggests that employers may be liable for discrimination based on a hidden disability once it is revealed through the hiring process. Given the invasive nature of data-based insights, a video interview algorithm may effectively disclose and penalize disability without an individual ever consenting to such disclosure, and this action would directly violate the ADA.

198. *Id.*

199. *Id.*

200. *EEOC v. Randstad*, No. 1:11-cv-01303 (D. Md. filed May 10, 2012).

201. *Randstad US Sued by EEOC for Disability Discrimination*, EEOC (May 13, 2011), <https://www.eeoc.gov/newsroom/randstad-us-sued-eeoc-disability-discrimination>.

202. *Id.*

203. *Id.*

204. *Id.*

205. *Randstad US, LP to Pay 60,000 to Settle EEOC Disability Bias Suit* (May 10, 2012), <https://www.eeoc.gov/newsroom/randstad-us-lp-pay-60000-settle-eeoc-disability-bias-suit>.

Although automated video interviewing is still a relatively new practice, there is some case law concerning the legality of personality testing under the ADA. This case law proves compelling, if not controlling, precedent for certain video interview algorithms which also test for personality traits. HireVue, for example, states its assessments are designed to produce “excellent insight into attributes like social intelligence (interpersonal skills), communication skills, *personality traits*, and overall job aptitude.”²⁰⁶ Given the vague yet potentially invasive nature of the insights video interview algorithms produce concerning an individual’s personality, it is valuable to consider how case law has treated personality testing under the ADA when evaluating protections for video interview candidates.

To the extent that automated video interviewing systems are also personality tests, the case of *Thompson v. Borg-Warner Protective Services Corp* (1996) has helped to establish these automated systems do not necessarily violate the ADA’s medical examination clause in all instances.²⁰⁷ The court found that plaintiff Bog-Warner’s use of a personality test called PASS-III to screen security guard applicants was legal.²⁰⁸ The court directly applied the factors laid out by the EEOC’s guidance to determine that PASS-III was not a medical exam for ADA purposes.²⁰⁹ By distinguishing between prohibited pre-offer medical exams and preemployment assessments that provided “information surrounding an applicant’s character or personality traits, and their fit for the job,” the court effectively found that personality tests may be permitted by the ADA in some forms.²¹⁰ This precedent means that video interview assessments must therefore meet more specific criteria to invoke the ADA’s medical examination protection.

The Seventh Circuit Court of Appeals decision in *Karraker v. Rent-A-Center* (2005) sheds insight on what an assessment that violates the ADA’s medical examination clause may look like. In *Karraker*, the court found that Rent-A-Center’s use of the Minnesota Multiphasic Personality Inventory (MMPI) as one of many variables in their pre-promotion test constituted discrimination under the ADA because “although applicant responses were not interpreted by a medical professional, the use of the MMPI would still be likely to identify

206. Patricia Barnes, *Artificial Intelligence Poses New Threat to Equal Employment Opportunity*, FORBES (Nov. 10, 2019, 1:57 PM) (emphasis added), <https://www.forbes.com/sites/patriciabarnes/2019/11/10/artificial-intelligence-poses-new-threat-to-equal-employment-opportunity/?sh=6e0a33036488>.

207. *Thompson v. Borg-Warner Protective Servs. Corp.*, No. C-94-4015, 1996 WL 162990 (N.D. Cal. Mar. 11, 1996).

208. *Id.* at *9.

209. Melson-Silimon et al., *supra* note 193, at 122–23.

210. *Id.* at 123.

and ‘weed out’ individuals with PDs who are protected under the ADA.”²¹¹ The court found that the MMPI was at least partly designed to identify mental illness and thus constituted a medical examination.²¹² Applied to video interview algorithms, this case shows that algorithms need not be interpreted by a doctor to violate the ADA; they need only be proven to be designed even in part to reveal mental impairments. Based on *Karraker’s* precedent, any video interview algorithm that incorporated the MMPI or a similar medical assessment in its design may violate the ADA. However, given the opaque nature of algorithms, proving such integration would be nearly impossible. Furthermore, the MMPI is a more obvious example of an assessment designed to reveal mental impairments, given its use as a medical diagnostic tool. It is not clear how courts would apply this precedent to proprietary algorithmic insights, which de facto reveal impairments by coding for particular traits that are proxies for disability.

Even if courts found that automated video interviewing constituted an illegal medical assessment under the ADA, job candidates may still struggle to prevail on their claims. In *Barnes v. Cochran*, the court found that a preemployment psychological evaluation violated the ADA’s ban on pre-offer medical evaluation, applying the EEOC’s seven factor guidance to the case.²¹³ However, the court nonetheless ruled in favor of the employers, reasoning that the plaintiff did not meet their burden of proof to show that “employment was denied for discriminatory reasons,” thus mooted the ADA violation.²¹⁴ According to Melson-Silimon and her coauthors, “[t]his decision highlights the burden plaintiffs face when suing on the grounds of disability-based discrimination; specifically, any legitimate justification articulated by the defendant for an adverse employment decision must be proven by the plaintiff to be a pretext for discrimination.”²¹⁵ Given that employment algorithms consider thousands of different data points, it may be nearly impossible to prove that the disability in question was the deciding factor in the algorithm’s ultimate employment recommendation. This is an issue not limited to medical examination cases, but one central to all ADA claims which may be brought against video interview algorithms.

211. *Id.* (quoting *Karraker v. Rent-A-Center, Inc.*, 411 F.3d 831, 837 (7th Cir. 2005)).

212. Abdi Tinwalla & J. Richard Ciccone, *ADA and Medical Examinations*, 34 J. AM. ACAD. PSYCHIATRY L. ONLINE 255, 256 (2006).

213. Melson-Silimon et al., *supra* note 193, at 123.

214. *Id.*

215. *Id.*

Critically, the ADA was modeled in part as parallel legislation to the Civil Rights Act of 1964.²¹⁶ Title I of the ADA specifically and intentionally mirrors Title VII, down to EEOC enforcement power granted over both statutes.²¹⁷ Applying the ADA to video interviewing thus faces many of the same challenges as seen in a Title VII case. ADA claims follow a similar litigation structure to Title VII claims, though they largely fall under disparate treatment theories rather than impact.²¹⁸ This means that an applicant would need to prove an employer would not have made the adverse employment decision in question but for the individual's disability. Thus, a claimant providing evidence that a video interview algorithm constituted a prohibited pre-offer medical examination or could have screened out candidates with disabilities is not enough.²¹⁹ The candidate must still prove that in their specific case: the causative reason for why they did not get a job offer was that the interviewing algorithm screened them out on the basis of a disability. Satisfying such a burden of proof would require a deep insight into the algorithm in question, a level of access which the job applicant would almost certainly be denied. Prevailing on an ADA claim would therefore prove a serious challenge in the face of the opaque nature of hiring algorithms given that (1) many applicants are prevented from examining the hiring algorithms, and (2) the black box nature of some algorithms makes it difficult to ascertain how exactly the discrimination happened.²²⁰

C. PRIVACY LAW PROTECTION FOR JOB APPLICANTS?

Besides potentially discriminatory harms, automated video interviewing systems pose great privacy risks because, as a necessary means to quantifying the veracity and character of job applicants, they capture a treasure trove of biometric data. Thus, another question is whether there are any extant privacy laws that can provide some legal protection to job applicants. The Sections below briefly review the different genres of privacy laws and policies and their limitations.

216. Robert D. Dinerstein, *The Americans with Disabilities Act of 1990: Progeny of the Civil Rights Act of 1964*, AM. BAR ASS'N (July 1, 2004), https://www.americanbar.org/groups/crsj/publications/human_rights_magazine_home/human_rights_vol31_2004/summer2004/irr_hr_summer04_disable/.

217. *Id.*

218. Third Circuit Model Jury Instruction for Employment Claims Under the Americans with Disabilities Act (Apr. 19, 2019), https://www.ca3.uscourts.gov/sites/ca3/files/9_Chap_9_2019_April.pdf.

219. *Id.*

220. Desai & Kroll, *supra* note 18, at 636.

1. *Notice and Consent*

In the United States, federal information privacy law and policy generally follows a framework known as “notice-and-consent.”²²¹ Legal scholar Daniel Susser explains the origins of this framework in *Notice After Notice-and-Consent*, as he examines common criticism of the policy, advocating for the importance of privacy disclosures despite concerns about consent.²²² Notice-and-consent grew out of a 1973 project by the U.S. Department of Health, Education, and Welfare (HEW) to mitigate “the threat to individual privacy posed by the government’s move toward computerized record-keeping.”²²³ HEW’s response was to establish the “‘Fair Information Practice Principles’ (FIPPs)” to guide regulation and policymaking around information privacy.²²⁴ Critically, the FIPPs are only guidance: they do not in and of themselves have the weight of law.²²⁵ Rather, they “encourage” compliance through the threat of “Federal Trade Commission (FTC) enforcement actions” on the basis “‘unfair and deceptive’ trade practices.”²²⁶ Thus, the regulatory value of the FIPPs heavily depends on how the FTC conceptualizes and enforces them. The FTC updated the FIPPs in 2000 “as guidance for designing commercial privacy policies.”²²⁷ The revised FIPPs offer four recommendations concerning “Notice,” “Choice,” “Access,” and “Security,” stating that:

1. Notice—Websites would be required to provide consumers clear and conspicuous notice of their information practices, including what information they collect, how they collect it (e.g., directly or through nonobvious means such as cookies), how they use it, how they provide choice, access, and security to consumers, whether they disclose the information collected to other entities, and whether other entities are collecting information through the site.
2. Choice—Websites would be required to offer consumers choices as to how their personal identifying information is

221. Daniel Susser, *Notice After Notice-and-Consent: Why Privacy Disclosures Are Valuable Even If Consent Frameworks Aren’t*, 9 J. INFO. POL’Y 37, 37 (2019).

222. *Id.*

223. *Id.* at 39.

224. *Id.* at 39–40.

225. See U.S. GOV’T ACCOUNTABILITY OFF., GAO-14-81, IN-CAR LOCATION-BASED SERVICES, APPENDIX I: OBJECTIVES, SCOPE, AND METHODOLOGY 23, 25 n.4 (2013) (“FIPPs are widely accepted principles for protecting the privacy and security of personal information. They were first proposed in 1973 by a U.S. government advisory committee. FIPPs are not precise legal requirements. Rather, they provide a framework of principles for balancing the need for privacy with other interests.”).

226. Susser, *supra* note 221, at 41.

227. *Id.*

used beyond the use for which the information was provided

3. Access—Websites would be required to offer consumers reasonable access to the information a website has collected about them, including a reasonable opportunity to review information and to correct inaccuracies or delete information.
4. Security—Websites would be required to take reasonable steps to protect the security of the information they collect from consumers.²²⁸

These principles gave rise to the notice-and-consent regime.²²⁹ Susser purports that the FIPP revisions are significant given their “procedural” nature. Because the FTC “drop[ped] the substantive concerns about data reliability and purpose specificity” that were central to the original FIPPs, the resulting notice-and-consent framework essentially allows employers to use consumer information as they see fit, so long as consumers knowingly agree.²³⁰ According to Susser, critics have panned this “free-market approach to privacy” for (1) not truly providing consumers “real options to choose from”; (2) allowing businesses to exploit “information asymmetries” at the expense of the uninformed consumer; (3) proving to be an unfeasible method for “engag[ing] with a huge number of information actors” in the modern day; (4) forcing consumers to “make onetime decisions” about particular pieces of data without knowing the long term “aggregate” effects of that data; and (5) ignoring the “social interests” inherent to data, instead vesting all decision-making authority with consumers.²³¹ Susser, following various other critics including Joel Reidenberg, Solon Barocas, and Helen Nissenbaum, ultimately declares “[n]otice-and-consent . . . to be a failed regulatory model.”²³² He joins other scholars in proposing an alternative model for regulating information privacy in the age of Big Data, explored in more detail below.

The notice-and-consent framework’s failures are especially salient in the context of algorithm-based video interviewing. Consenting to give up one’s data rights in the video interview process may not feel like much of a choice when employment is at stake; companies may not offer, or advertise that they offer, any meaningful alternative method of job candidate evaluation. As such, the nature of the hiring process means candidates may consent by default. Furthermore, it is important to consider when and how employers provide

228. *Id.* at 41–42.

229. *Id.* at 42.

230. *Id.* at 41.

231. *Id.* at 42–46.

232. *Id.* at 43–47.

notice disclosures to candidates. Some video interview vendors only act as “data processer[s]”; that is, *employers* retain the rights to control the data candidates provide, not the software company itself.²³³ Thus, a candidate cannot simply turn to a vendor’s website to understand how their data will be used. They instead must seek out an employer’s privacy policy directly.

These practices pose two potential issues. First, how employers choose to provide a privacy notice would likely have a big impact on whether the candidate was actually capable of consenting: if the disclosure occurred right before a candidate started an interview, it is likely that the applicant may see consent as part of the bargain to have the opportunity for an interview. Second, delegating data control to employers means that a candidate’s privacy rights are directly tied to the power asymmetry of the preemployment relationship. Candidates may be less likely to ask questions or request data access from an employer for fear of risking their job opportunity. What’s more, candidates who engage with the same video interview software for interviews across multiple companies may not realize that their privacy rights are changing with each successive interview. As such, they may only read the first disclosure and consent to all successive disclosures under the assumption that the substance is the same. This potential confusion is significant given the serious privacy issues inherent to video interviewing, discussed in more detail below. Above all, as Susser identified, notice-and-consent’s procedural protections do not address any of the substantive privacy issues that candidates face. If an employer chooses to share the highly sensitive, aggregated data insights they mined from a candidate’s interview with other businesses or potential employers, it’s not clear what substantive right notice-and-consent would give a candidate over their data if the candidate had already signed an initial, broad consent agreement.

2. *State Law*

Given the massive gaps federal privacy law leaves, some states have taken steps to protect against the threat of employers harnessing the power of Big Data. For example, in 2019, Illinois passed the Artificial Intelligence Video Interview Act (AIVIA), specifically designed to govern privacy risks associated

233. *HireVue Privacy Notice*, HIREVUE, <https://www.hirevue.com/privacy#what-info-does-hirevue-collect> (Jan. 20, 2021) (“If you are a job candidate (“Candidate”) or employee (“Employee”) using our Services on behalf of one of our customers who are engaging us to provide the Services to them (the ‘Potential Employers’), we are collecting and processing your personal information on behalf of the Potential Employers. In such cases, we are acting as a data processor and are collecting and processing your personal information on their behalf and in accordance with their instructions.”).

with video interview assessments.²³⁴ This law, dubbed “the first of its kind in the US,”²³⁵ includes five main requirements to which employers using AI video technology, such as HireVue, must adhere. First, employers are required to “[n]otify the applicant, in advance, that the organization is using the technology to analyze video interviews.”²³⁶ The law further mandates that employers “[e]xplain to the applicant ‘how the [AI] works’ and what general characteristics the technology uses to evaluate applicants.”²³⁷ This clear call for transparency is helpful. However, many video technology companies do not publish adequate information on the workings of their products.²³⁸ Thus, the effects of this part of the law may take one of two paths: either AI video providers will be forced to publish more information about their algorithms *or* the standard for meeting this transparency mandate will be effectively so low as to render it meaningless. Beyond transparency, the law requires that employers “[o]btain, in advance, the applicant’s consent to use the technology.”²³⁹ The law also features provisions for data protection. It imposes limits on “the distribution and sharing of the video,” granting access “to only those persons ‘whose expertise or technology’ is necessary to evaluate the applicant.”²⁴⁰ Further, candidates are given some control over what happens to the video after their assessment. Employers are required to “destroy the video (and all backup copies) within 30 days” of the applicant requesting its destruction.²⁴¹

Law firm David Wright Tremaine LLP (DWT) identifies a few key issues with the law. Chiefly, the law fails to define “‘artificial intelligence’ and ‘artificial intelligence analysis’ ” along with other “key terms.”²⁴² This ambiguity may mean that certain employer AI uses, such as “to track data about its candidates,” may not be covered.²⁴³ Further, ambiguity in the transparency part

234. Rebecca Heilweil, *Illinois Says You Should Know If AI Is Grading Your Online Job Interviews*, VOX (Jan. 1, 2020), <https://www.vox.com/recode/2020/1/1/21043000/artificial-intelligence-job-applications-illinois-video-interview-act>.

235. *Id.*

236. Matthew Jedreski, Jeffrey S. Bosley & K.C. Halm, *Illinois Becomes First State to Regulate Employers’ Use of Artificial Intelligence to Evaluate Video Interviews*, DAVIS WRIGHT TREMAINE LLP (Sept. 3, 2019), <https://www.dwt.com/blogs/artificial-intelligence-law-advisor/2019/09/illinois-becomes-first-state-to-regulate-employers>.

237. *Id.*

238. *See generally* Desai & Kroll, *supra* note 18, at 636 (arguing that many algorithmic systems are “black box” systems with little explanation of their workings).

239. Jedreski et al., *supra* note 236.

240. *Id.*

241. *Id.*

242. *Id.*

243. *Id.*

of the law may, as suggested above, poses serious problems for its effective use. DWT notes that the law does not go in-depth to specify or define “how much detail about the AI technology an employer must provide when ‘explaining how artificial intelligence works’ to an applicant” or what “ ‘characteristics’ of the AI employers must disclose.”²⁴⁴ Therefore, employers may be permitted to use broad, cursory statements such as “AI will assess a candidate’s performance”²⁴⁵ to satisfy this requirement—statements that do not serve the true spirit of transparency. There is further no requirement that candidate consent be expressly written.²⁴⁶ DWT notes, further, that the law “does not include a private right of action or any explicit penalties”; this could raise serious issues in enforcing its provisions.²⁴⁷ As for data destruction, DWT points out that it is not clear if “data that an employer extracts or derives from the video interviews . . . is subject to the destruction duty under the law.”²⁴⁸ If such data is not protected by AIVIA, then the extent to which the act allows candidates control over their interview data is potentially limited. Lastly, DWT points out that “there is no guidance on what it means for a job to be ‘based in’ Illinois, and the statute is silent as to whether employees may refuse to consider applicants who refuse to consent.”²⁴⁹

Ultimately, AIVIA is a step in the right direction, as it touches on the serious concerns of transparency and data rights. However, the primary, overarching issue with the act is a lack of specificity. Failing to define key terms, expand on essential provisions, or stipulate any enforcement mechanism means that the effective impact of transparency and data rights measures is limited and employers who wish to evade the law may do so. Further, although some employers may surely make a good faith effort to comply, many employers themselves are not privy to how the AI they use truly works. Companies such as HireVue keep a close guard over their algorithms and technologies to protect their market share, to the detriment of clients and candidates alike. In order to push AI video interview companies to be more transparent, the law must put in place effective penalties such that employers would not choose to use technology unless AI companies provided enough information. Effective legislation must hold enough weight to impact all stakeholders in the AI video interview universe. Again, it is important to reiterate that Illinois is “at the forefront of regulating technology and personal

244. *Id.*

245. Ajunwa, *supra* note 7, at 644.

246. Jedreski et al., *supra* note 236.

247. *Id.*

248. *Id.*

249. *Id.*

data.”²⁵⁰ AIVIA should be commended as first-of-its-kind legislation that is shedding light on critical issues of public interest. It simply needs to go further to counterbalance the immense power that the AI sphere currently holds. Regardless, AIVIA acts as a model for other states to specifically protect consent and disclosure data rights around video interviewing. Given that federal protections may not apply, such specific legislation is an important first step to protecting applicant data.

Another Illinois law, the Biometric Information Privacy Act (BIPA), passed in 2008, offers more substantive protections around the specific issue of biometric privacy.²⁵¹ Key BIPA provisions around biometric data collection and use by businesses include “informed consent,” “a limited right to disclosure,” “protection obligations and retention guidelines,” “prohibit[ions on] profiting from biometric data,” “a private right of action for individuals harmed by BIPA violations,” and provisions for “statutory damages.”²⁵² Given that video interview assessments varyingly consider vocal and facial expressions, assessments may actually qualify for BIPA protections as biometric data refers to “the measurement and statistical analysis of an individual’s physical and behavioral characteristics,” including “voice prints,” “face . . . features,” “gestures,” and “voice.”²⁵³ While BIPA is primarily procedural in nature—again adhering to the federal notice-and-consent framework—it does afford candidates the right to sue and protections concerning third party access to sensitive biometric data. This is important considering the serious potential harm that may come to candidates if sensitive biometric interview data is sold to third parties, not in the least limited to the threat of deep fakes as discussed in a later Section. BIPA therefore fills a gap as it encodes specific kinds of information privacy in law, though it stops short of prohibiting the collection of such information altogether. Unfortunately, while other states, including Texas and Washington, have passed similar laws, these states appear to offer even more limited protections than Illinois.²⁵⁴

250. *Id.*

251. JACKSON LEWIS, ILLINOIS BIOMETRIC PRIVACY ACT FAQs (2021), <https://www.jacksonlewis.com/sites/default/files/docs/IllinoisBIPAFaqs.pdf>.

252. *Id.*

253. *Id.*

254. *See Collection of Biometric Data Raises Privacy Concerns for Employees and Compliance Issues for Employers*, FISHER PHILLIPS (Mar. 15, 2018), <https://www.fisherphillips.com/Employment-Privacy-Blog/collection-of-biometric-data-raises-privacy-concerns>; *see also* Capture of Use of Biometric Identifier Act, 50 TEX. BUS. & COM. CODE ANN. § 503.001 (resembling BIPA by requiring that, prior to being authorized to collect biometric identifiers: (1) the organization must obtain informed consent that (2) need not be in writing, (3) from individuals; but, differing from Illinois’ state law by only allowing the Texas Attorney General

Therefore, while offering a partially useful model, BIPA does not constitute or represent sweeping biometric privacy protections at the state level.

Some states have gone beyond specific privacy applications, instead creating more broad privacy protections to govern information exchanges at large. The California Consumer Privacy Rights Act (CCPA) offers one such example.²⁵⁵ Recently passed in 2020, CCPA gives consumers specific, enumerated rights over their data including the

1. Right to Correct Inaccurate Information . . .
2. Right to Have Personal Information Collected Subject to Data Minimization and Purpose Limitations . . .
3. Right to Receive Notice from Businesses Planning on Using Sensitive Personal Information and Ask Them to Stop . . .
4. Right to Access Information . . . [and]
5. Right to Opt Out of Sharing Information with Third Parties.²⁵⁶

As of January 1, 2021, CCPA protections were extended to Californian job applicants.²⁵⁷ Although the CCPA largely follows notice-and-consent frameworks, it takes significant steps towards giving consumers and employees meaningful control over their data by allowing individuals to opt-out of data sharing and certain uses of their data over its lifespan.

Given the law's newness, it's hard to measure its practical effects; reports suggest that the law's launch has resulted in a mix of "firms . . . disclosing too little data—or far too much."²⁵⁸ Companies such as Uber and Lyft have been selective as to what data they choose to disclose and what they choose to

to enforce the law as the law does not provide a private right of action); H.B. 1493, 65th Leg., 2017 Sess. (Wash. 2017) (limiting the definition of "biometric data" so that it likely excludes the facial recognition technology social media and photo storage websites use to automatically tag users in digital photographs and applying the law only to those biometric identifiers who are "enrolled" in a commercial database).

255. *California Privacy Rights Act*, PRIV. RTS. CLEARINGHOUSE (Dec. 10, 2020), <https://privacyrights.org/resources/california-privacy-rights-act-overview#:~:text=The%20California%20Privacy%20Rights%20Act%20clarifies%20that%20people%20can%20opt,per%20sonal%20information%20to%20third%20parties.&text=The%20California%20Privacy%20Rights%20Act%20expands%20this%20to%20cover%20data,includes%20a%20username%20and%20password.>

256. *Id.*

257. California Consumer Privacy Act of 2018, A.B.-25, 2019–2020 Sess. (Cal. 2019), https://leginfo.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB25.

258. Greg Bensinger, *So Far, Under California's New Privacy Law, Firms Are Disclosing Too Little Data—Or Far Too Much*, WASH. POST (Jan. 21, 2020, 7:44PM), www.washingtonpost.com/technology/2020/01/21/ccpa-transparency/.

retain.²⁵⁹ One Los Angeles man who tried to access his data reported that “[e]veryone seems to be . . . see[ing] what they can get away with I hate to say it, but I think the companies are going to win.”²⁶⁰ Thus, compliance remains a point of contention. Even if a state creates an all-encompassing information privacy law that extends to consumers and job applicants alike, ensuring that companies actually comply with the law is a massive regulatory task that state level agencies may struggle to keep up with. This reality makes the need for federal regulation with comprehensive enforcement mechanisms all the more critical.

On the whole, state laws offer some information privacy protections for certain states’ citizens who fall within certain categories. However, essentially no federal or state law offers an affirmative declaration of the data rights of job applicants. Notice-and-consent guidance has resulted in a serious gap in substantive protections. These patchwork state protections ultimately do not provide comprehensive protections.

3. *Fair Credit Report Act (FCRA) to the Rescue?*

The Fair Credit Reporting ACT (FCRA) is a “1970 [law enacted] to regulate the credit reporting industry because of concerns about the fairness and accuracy of credit reports.”²⁶¹ In recent years, legal scholars, and even the FTC, have suggested that its consumer privacy protections may extend to businesses using consumer data and data-based insights.²⁶² Thus, it is important to consider what, if any, privacy protections the FCRA may offer to video interview candidates.

The FCRA governs “compan[ies] . . . collecting and sharing third-party data that is used or expected to be used as a factor in determining eligibility for credit, insurance, employment, or other purpose[s] authorized under the

259. *Id.*

260. *Id.*

261. Pauline T. Kim & Erika Hanson, *People Analytics and the Regulation of Information Under the Fair Credit Reporting Act*, 61 ST. LOUIS U. L.J. 17, 20 (2016).

262. *See id.* at 28; *see also* Ajunwa, *supra* note 7, at 655; Karen Sanzaro, *Big Data: FTC Issues Report Cautioning that Use of Big Data May Violate Federal Consumer Protection Laws or Raise Ethical Considerations*, ALSTON & BIRD: PRIV., CYBER, & DATA STRATEGY BLOG (Jan. 19, 2016), <https://www.alstonprivacy.com/big-data-ftc-issues-report-cautioning-that-use-of-big-data-may-violate-federal-consumer-protection-laws-or-raise-ethical-considerations/> (summarizing FTC warning that companies using Big Data may be subject to the FCRA, references FTC enforcement actions against a firm that used consumer data for “eligibility determinations” without complying to FCRA).

FCRA.”²⁶³ These companies are considered “consumer reporting agencies” (CRAs) under the FCRA, formally defined as

any person which, for monetary fees, dues, or on a cooperative nonprofit basis, regularly engages in whole or in part in the practice of assembling or evaluating consumer credit information or other information on consumers for the purpose of furnishing consumer reports to third parties.²⁶⁴

Legal scholars Pauline T. Kim and Erika Hanson note that “entities that assemble and evaluate information for noncommercial uses as well as entities that assemble information about the entity’s own interactions with its customers” are not considered CRAs.²⁶⁵ Therefore, employers likely could not qualify as CRAs as interview reports would be for internal, noncommercial use; however, external video interview vendors who provide assessments to employers may. Thus, from the outset, it seems that the FCRA may govern video interview vendors to the extent that the data collected during a video interview is (1) for commercial use, and (2) considered a consumer report. Kim and Hanson refer to a three-prong framework that courts have developed to determine if “information constitutes a consumer report under the law”:

1) the information was communicated by the consumer reporting agency; 2) it bears on the “consumer’s credit worthiness, character, general reputation, personal characteristics, or mode of living”; and 3) it was “used or expected to be used or collected in whole or in part for one of the enumerated purposes.”²⁶⁶

All “elements” must be “satisfie[d]” to constitute a consumer report.²⁶⁷ Also expressly excluded from “consumer reports” are “report[s] containing information solely as to transactions or experiences between the consumer and the person making the report.”²⁶⁸ It seems plausible that video interviews may fall within this exclusion: the only consumer-specific data that interview assessments consider is collected from the interaction between the candidate and the algorithm. However, the algorithms *do* consider thousands of external

263. Chi Chi Wu, *Data Gatherers Evading the FCRA May Find Themselves Still in Hot Water*, NAT’L CONSUMER L. CTR. (June 14, 2019), <https://library.nclc.org/data-gatherers-evading-fcra-may-find-themselves-still-hot-water>.

264. *Id.* (quoting 15 U.S.C. § 1681a(f)).

265. Kim & Hanson, *supra* note 261, at 21–22.

266. *Id.* at 22 (quoting *Ernst v. Dish Network, LLC*, 49 F. Supp. 3d 377, 381 (S.D.N.Y. 2014) (citing cases from the U.S. Courts of Appeals)).

267. *Id.*

268. 15 U.S.C. § 1681a(d).

data points about other individuals.²⁶⁹ Although this is not information about the consumer, it is information used to make judgments and assumptions about the consumer which are not limited to the “transactions or experiences between the consumer” and reporter.²⁷⁰ The question would be to what extent this external information is actually “contain[ed]” within the report.²⁷¹

Thus, it seems possible that video interviews, where vendors collect candidate data to determine a candidate’s “character” or “personal characteristics” (amongst other things) for the purposes of employment eligibility qualify, could qualify as consumer reports under the FCRA.²⁷² Therefore, video interview vendors would likely qualify as CRAs. As I explored in a prior law review article, *The Paradox of Automation as Anti-Bias Intervention*, applying FCRA frameworks to hiring algorithms “may...enable the job applicant to discover if the employer had access to discriminatory information or even to establish a pattern of discriminatory information furnished to the employer for protected groups, thus perhaps assisting in a disparate impact cause of action.”²⁷³ As a CRA, vendors would be required to “follow reasonable procedures to assure the maximum possible accuracy of [their] files,”²⁷⁴ including allowing “consumers to review information in their files without charge, investigat[e] alleged inaccuracies, and provid[e] information to consumers about their rights.”²⁷⁵ Employers, as the entity using the consumer report, would be required to

provide a clear, conspicuous, and stand-alone disclosure [to applicants] that a consumer report may be obtained for employment purposes; they would be required to request written authorization from the applicant or employee for procurement of the report; and certify to the consumer reporting agency its compliance with the requirements of the statute and that it will not violate any equal employment opportunity law.²⁷⁶

Furthermore, the FCRA would require that an employer “provide notice before rejecting a job application . . . or taking any other adverse employment action” in addition to “provid[ing the applicant] a copy of the consumer report relied upon and a description of the individual’s rights under the FCRA,”

269. See *supra* Part II(b)(3) (denoting that external information is needed for a report to be considered a consumer report).

270. 15 U.S.C. § 1681(a)(d)(2)(i).

271. *Id.*

272. *Id.* at § 1681a(e).

273. Ajunwa, *supra* note 8, at 1735.

274. *Id.* at 1740.

275. Kim & Hanson, *supra* note 261, at 22–23.

276. *Id.* at 23.

which include “an opportunity to review the report and attempt to correct any mistakes.”²⁷⁷ After rejecting the applicant, the employer would further have to follow through with several more procedural steps, including providing information about the CRA who provided the report and “notice of the individual’s rights to dispute the accuracy or completeness of the report and to receive an additional copy of the report if requested within sixty days.”²⁷⁸ Failure to comply would result in FTC enforcement action.²⁷⁹

As Kim and Hanson note, the FCRA’s protections are “procedural.”²⁸⁰ Indeed, the FCRA does not offer job applicants any substantive right to privacy and does not “[limit] . . . the *types* of information that can be collected or reported.”²⁸¹ However, if video interviews were considered consumer reports under the FCRA, it seems possible that FCRA protections may ameliorate some problems inherent to video interviewing. Particularly, given the opaque nature of algorithms, disclosures concerning the reasoning for an adverse employment action on the basis of the interview may provide valuable “insight as to how [candidates] are evaluated” and could help society “regain some measure of checks over the information that is used to ‘screen’ candidates as part of the automated hiring trend.”²⁸²

Of course, these protections do not go far enough to control what kind of invasive data employers collect and how they use it. As Spencer Mainka observes in *Algorithm-Based Recruiting Technology in the Workplace*, “[t]he FCRA provides no relief for an applicant who was denied an opportunity based on inaccurate data because the FCRA only regulates the process.”²⁸³ In this way, FCRA follows the same pattern of free-market regulation as notice-and-consent. The invasive nature of the privacy threats that video interviewing poses requires more substantive protections. Beyond all of this, video interviews are likely excluded from FCRA protection, falling within the exclusion of “report[s] containing information solely as to transactions or experiences between the consumer and the person making the report.”²⁸⁴ The FCRA’s privacy protections may therefore not even apply at all. Regardless, it is useful to consider the utility of disclosure, central to the FCRA’s

277. *Id.*

278. *Id.* at 24.

279. *Id.*

280. *Id.*

281. *Id.* at 25.

282. Ajunwa, *supra* note 8, at 1741.

283. Spencer Mainka, *Algorithm-Based Recruiting Technology in the Workplace*, 5 TEX. A&M J. PROP. L. 801, 815 (2019).

284. 15 U.S.C. § 1681a(d).

frameworks, in combatting the opaque nature of algorithmic decision-making for employment.

IV. APPLYING A LEX INFORMATICA FRAMEWORK

Given the identified limitations of existing law to address the unlawfully discriminatory potential of automated video interviewing, it is important to consider other types of regulatory frameworks. In his prescient 1998 article, *Lex Informatica*, legal scholar Joel R. Reidenberg identified “three substantive legal policy areas” that he argued were “in a critical state of flux in the network environment,” similar to the instability early merchants faced as they navigated jurisdictions.²⁸⁵ Those areas were: “[t]he treatment of content, the treatment of personal information, and the preservation of ownership rights.”²⁸⁶ For merchants, the solution came in the form of “a distinct body of law known as the ‘Lex Mercatoria.’” Influenced by “[c]ustom and practices” of the trade, Lex Mercatoria acted “independent of local rules and assured commercial participants of basic fairness in their relationships.”²⁸⁷ Reidenberg argues that the rules of information technology can act in the same way. That is, in order to properly regulate information flows in the digital age, policymakers must first turn to “the set of rules for information flows imposed by technology and communication networks” which foster their own “Lex Informatica.”²⁸⁸

A true benefit of a Lex Informatica framework is that it “relies typically on *ex ante* measures of self-execution.”²⁸⁹ Unlike the current aspects of the U.S. legal regime, which some scholars have criticized as “backward-looking,”²⁹⁰ Lex Informatica “allows automated monitoring of information access and use,” preventing rule violations from ever occurring.²⁹¹ Whereas traditional law requires candidates to know a violation of their rights occurred in order to seek protection—a serious problem given the opaque nature of algorithmic decision-making—technological solutions under a Lex Informatica framework provide some assurance that such violations will not occur in the first place by, for one, addressing design elements that aid in discriminatory practices. This *ex ante* aspect is especially valuable given the permanent harms impermissible data use may inflict on candidates; just as spilled milk can never be fully

285. Reidenberg, *supra* note 13, at 554.

286. *Id.*

287. *Id.* at 553.

288. *Id.* at 554–55.

289. *Id.* at 581 (emphasis added).

290. Kim, *supra* note 7, at 867–68 (“Addressing the challenges of workforce analytics using a theory of classification bias also reveals the limitations of the backward-looking, liability-focused model of legal regulation embodied by Title VII.”).

291. Reidenberg, *supra* note 13, at 581.

returned to the carton, exposed data can never be fully recovered and protected, no matter what a court orders.

Applying a Lex Informatica framework to video interviewing means developing legislation that considers the capabilities of the technology itself rather than solely how the actors intend to use it or the use in practice. It is important to underscore here that Lex Informatica is not techno-solutionism. As Reidenberg emphasizes, rather than a replacement for all traditional regulation, “Lex Informatica must be seen as a distinct source of policy action. Effective channeling of Lex Informatica requires a shift in the focus of government action away from direct regulation and toward indirect influence.”²⁹² Reidenberg chiefly argues that government must seek to influence how technology is developed from its inception, thus impacting technological design and development by participating in “funding” to “regulate[] behavior and . . . standards,” rather than merely seeking to address just the consequences of technology.²⁹³

A. TREATMENT OF CONTENT

Applying a Lex Informatica framework to the video interviewing process also means considering the treatment of content acquired from candidates. I concur with legal scholars who have argued that the Uniform Guidelines on Employee Selection Procedures²⁹⁴ should apply in negotiating what content will be digested by automated hiring systems.²⁹⁵ Although the Uniform Guidelines are not law,²⁹⁶ they are seen as authoritative²⁹⁷ and have influenced decisions in employment discrimination cases.²⁹⁸

292. *Id.* at 586.

293. *See id.* at 588.

294. Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607 (2021).

295. *See Sullivan, supra* note 7, at 420–22.

296. *Id.* at 422.

297. *See Griggs v. Duke Power Co.*, 401 U.S. 424, 433–34 (1971) (concluding that the EEOC’s interpretation of the guidelines should be given “great deference”); *see also Albemarle Paper Co. v. Moody*, 422 U.S. 405, 430–31 (1975) (observing that the “Guidelines draw upon and make reference to professional standards of test validation established by the American Psychological Association,” and that while the guidelines were “not administrative ‘regulations’ promulgated pursuant to formal procedures established by the Congress . . . they do constitute [t]he administrative interpretation of the Act by the enforcing agency”); *Gulino v. N.Y. State Educ. Dep’t*, 460 F.3d 361, 384 (2d Cir. 2006) (discussing how in 1978 the Uniform Guidelines replaced the original EEOC guidelines and has since become the primary, authoritative “yardstick by which we measure defendants’ attempt to validate [a standardized certification test]”).

298. *Sullivan, supra* note 7, at 422 n.106 (noting that per the results of a Lexis Advance search on December 10, 2017, “[t]he Guidelines have been cited in more than 300 cases, including a number of Supreme Court decisions.”).

The Uniform Guidelines are useful because they set standards for when selection criteria could be considered valid. Thus, the Guidelines provide for “three kinds of validation: criterion, content and construct.”²⁹⁹ The aim of all three types of validation is to prompt the employer to provide evidence of a predictive causal relationship between the selection method and the job performance:

Evidence of the validity of a test or other selection procedure by a criterion-related validity study should consist of empirical data demonstrating that the selection procedure is predictive of or significantly correlated with important elements of job performance. Evidence of the validity of a test or other selection procedure by a content validity study should consist of data showing that the content of the selection procedure is representative of important aspects of performance on the job for which the candidates are to be evaluated. Evidence of the validity of a test or other selection procedure through a construct validity study should consist of data showing that the procedure measures the degree to which candidates have identifiable characteristics which have been determined to be important in successful performance in the job for which the candidates are to be evaluated.³⁰⁰

I thus interpret the Uniform Selection Guidelines as requiring that: (1) the variables used by the automated video interviewing algorithm relate to important aspects of the job, (2) the data from the automated video allow for the prediction of future job performance, and (3) the selected candidates from the automated video interview have characteristics that can be identified as casually linked to superior job performance.

1. *Criterion Validity for Automated Video Interviewing*

The current iteration of automated video interviewing systems fails criterion validity. The Uniform Guidelines requires for criterion validity: “Evidence of the validity of a test or other selection procedure by a criterion-related validity study should consist of empirical data demonstrating that the selection procedure is predictive of or significantly correlated with important elements of job performance.”³⁰¹ However, as discussed earlier in the Article,³⁰² experts dispute whether empirical evidence supports the conclusion that automated video interviewing is predictive of important elements of job

299. *Id.* at 423 (citing RAMONA L. PAETZOLD & STEVEN L. WILLBORN, *THE STATISTICS OF DISCRIMINATION* §§ 5.13–.17 (2d ed. 2017–2018)).

300. *Id.* (citations omitted) (quoting 29 C.F.R. § 1607.5B (2018)).

301. Uniform Guidelines on Employee Selection Procedures, 41 C.F.R. § 60–3.5 (2021).

302. *See* Ajunwa, *supra* note 7, at 642–43.

performance, such as a “veracity” or “conscientiousness.”³⁰³ Thus, since it is disputed that automated video interviewing systems can measure these variables, these systems have not met the standard for criterion validity.

2. *Content Validity for Automated Video Interviewing*

Automated video interviewing systems are also on shaky ground when it comes to content validity. The Uniform Guidelines maintain: “Evidence of the validity of a test or other selection procedure by a content validity study should consist of data showing that the content of the selection procedure is representative of important aspects of performance on the job for which the candidates are to be evaluated.”³⁰⁴ Scholars have described automated video interviewing systems as attempting to decipher a wide range of behaviors and personality states.³⁰⁵ This invites criticism that some of the variables that the automated hiring system is attempting to capture are simply not representative of important parts of the job that candidate is seeking.

3. *Construct Validity for Automated Video Interviewing*

The construct validity of automated video interviewing systems also seems uncertain. The Uniform Guidelines declare:

Evidence of the validity of a test or other selection procedure through a construct validity study should consist of data showing that the procedure measures the degree to which candidates have identifiable characteristics which have been determined to be important in successful performance in the job for which the candidates are to be evaluated.³⁰⁶

This part gets into the error rates of automated video interviewing systems. Even if these systems have been programmed to be predictive of job performance and the variables used do represent important aspects of the job, questions remain about whether the programs work accurately. The black box nature of many automated decision-making systems makes answering these questions difficult. An audit in which non-selected candidates are compared to

303. See Ajunwa, *supra* note 7, at 663, 677, 685.

304. 41 C.F.R. § 60–3.5(B).

305. See Kate Crawford, *Artificial Intelligence Is Misreading Human Emotion*, ATLANTIC (Apr. 27, 2021), <https://www.theatlantic.com/technology/archive/2021/04/artificial-intelligence-misreading-human-emotion/618696/>.

306. 41 C.F.R. § 60–3.5(B).

selected candidates (in a longitudinal study) would be one way to confirm the accuracy of automated hiring platforms.³⁰⁷

B. TREATMENT OF PERSONAL INFORMATION

Lex Informatica provides a regulatory framework for the vast trove of biometric data subsumed in the automated interviewing process. Much of these data reveal demographic characteristics and are also personally identifiable information (PII) and are thus highly sensitive information that should enjoy heightened legal protection. One advantage of the Lex Informatica framework is that it recognizes the role that the technological capabilities of technological systems could play in regulation. As I previously discussed in another law review article, one great technological capability is that decision-makers could segregate demographic data prior to an employment decision.³⁰⁸ In the context of automated video interviewing, this could hide the video from the human decision-maker behind an information wall and only share the scores from the interviewing algorithm.

The technological capabilities of an automated hiring system also provide other mechanisms for protecting PII. For example, the system could be designed to work on access keys, which allow certain parties to view the information or that restrict access after a period of time. In addition, the data containing PII could be programmed to self-destruct after a period of time or if unauthorized access is attempted.

C. PRESERVATION OF OWNERSHIP RIGHTS: “PRIVACY AS TRADE SECRECY”

Other types of enforcement systems such as a property rights enforcement model for the data of automated video interview candidates could also be feasible under the Lex Informatica framework, especially under the prong that calls for the preservation of ownership rights. In *Privacy as Intellectual Property?*, Professor Pamela Samuelson considers that, chiefly, “a property rights model” of information privacy would both “establish a right in individuals to sell their personal data and thereby capture some of the value their data have in the

307. See Ajunwa, *supra* note 7, at 672 (citing O’NEILL RISK CONSULTING AND ALGORITHMIC AUDITING, DESCRIPTION OF ALGORITHMIC AUDIT: PRE-BUILT ASSESSMENTS 1, 1–2 (2020), <https://webapi.hirevue.com/wp-content/uploads/2021/01/oneil-risk-consulting-and-algorithmic-auditing-01-2021.pdf> (“On January 11, 2021, HireVue announced that it had brought in the auditing entity, O’Neil Risk Consulting and Algorithmic Auditing (‘ORCAA’), to conduct an audit of its video its algorithms considered. The report of the audit, however, left many questions unanswered. For one, ORCAA limited the audit to ‘pre-built assessments used in hiring early career candidates, including from college campuses.’”).

308. See Ajunwa, *supra* note 7, at 651.

marketplace” as well as “force companies to internalize certain social costs of the widespread collection and use of personal data now borne by others.”³⁰⁹ However, she acknowledges that property interests and privacy interests do not always align, as “individuals may not have just one interest in personal information, but many interests,” which may differ according to individuals or circumstances.³¹⁰

Therefore, Professor Samuelson suggests “an alternative market-oriented legal regime for protecting personal information” built on “a default rule providing individuals with certain rights to control the collection or processing of personal information about them while also providing individuals with the power to contract away this right (e.g., when they receive compensation for doing so).”³¹¹ Samuelson proposes that trade secrecy law, as opposed to intellectual property, offers a buildable model to start from, as it “facilitates license transactions in information, while . . . providing default rules to govern uses and disclosures of protected information, and setting minimum standards of acceptable commercial practice.”³¹² Essentially, if information privacy policy was rooted in trade secrecy, individuals would have the power to “license” rights to their data for limited uses in particular circumstances without running the risk that third parties data uses or disclosures violate the parties’ agreement.³¹³

Identifying and enforcing a property right in the data collected from video interview job applicants could be part of a Lex Informatica regulatory approach for automated video interviewing. A start is to formally identify the job applicant’s property right in the biometric data collected as part of automated video interviewing. This could be supervised by the EEOC as the employment standards regulatory agency³¹⁴ or under the FTC, which considers lawful uses for technology used as part of commerce.³¹⁵ Establishing such a property right would mean that the job applicant could retain control over what data could be acquired by the employer and that the job applicant could also constrain the use cases for the data. Exercising control of the data could mean that the job candidate signs a pre-interview contract that serves as a license to the biometric data for the prospective employer. Not only would

309. Pamela Samuelson, *Privacy as Intellectual Property?*, 52 STAN. L. REV. 1125, 1128–29 (2000).

310. *Id.* at 1171–72.

311. *Id.* at 1129.

312. *Id.* at 1152.

313. *See id.* at 1155.

314. *See* Ajunwa, *supra* note 7, at 667.

315. *See, e.g.,* Frank Pasquale, *Beyond Innovation and Competition: The Need for Qualified Transparency in Internet Intermediaries*, 104 NW. U. L. REV. 105, 155 (2010).

such a license explicitly outline the types of data that could be collected from the job applicant, but the license could also delineate the boundaries for use cases of the data.

To illustrate, in another law review article, I have detailed the risk of “algorithmic blackballing” that could arise from the unfettered access to the applicant data that automated hiring currently affords the employer.³¹⁶ This means that the applicant data collected as part of an interview for a failed employment bid could be resurrected during a second bid for employment and once again used to thwart employment. A regulatory regime wherein job applicants license interview data in exchange for consideration for, and solely for, the purpose of being considered for a specific job position would eliminate much of the danger of algorithmic blackballing. This is because the applicant could, through licensing, retain control over the shelf life of the data collected and could dictate a hard delete of the data after the first job attempt.

V. CONCLUSION

Humans have long judged each other by physical appearance.³¹⁷ From time to time, there have been efforts to elevate this practice into science.³¹⁸ Yet, at each instance, the scientific method has revealed no clear causative link between a person’s facial features, facial expressions, and their character.³¹⁹ With automated video interviewing, we see an attempt to routinize this human practice as a matter of business procedure, to quantify the practice of judging the character of humans by physical traits, and to delegate this practice to machines. If finding the right employee can be likened to the romantic selection process, then the use of automated video interviewing and concomitant facial analysis may be likened to the myth of Narcissus. As the legend goes, Narcissus rejected many romantic prospects and instead fell in love with his own reflection. While admiring his reflection in a pool, Narcissus fell in and drowned. Automated hiring systems may be seen as mirrors that reflect to us the racial, gender, and ableist biases present in our society, biases which dictate who would make the ideal employee. The bedrock legal principle

316. See Ajunwa, *supra* note 7, at 622–23.

317. See Pam Belluck, *Yes, Looks Do Matter*, N.Y. TIMES (Apr. 24, 2009), <https://www.nytimes.com/2009/04/26/fashion/26looks.html> (discussing the historical and modern role of physical appearance in human decision-making).

318. See Matt Simon, *Fantastically Wrong: The Silly Theory That Almost Kept Darwin From Going On His Famous Voyage*, WIRED (Jan. 21, 2015, 6:30 AM), <https://www.wired.com/2015/01/fantastically-wrong-physiognomy/> (discussing the long history of physiognomy, a pseudo-scientific field which purported one’s facial features betrayed their character).

319. See Sahil Chinoy, *The Racist History Behind Facial Recognition*, N.Y. TIMES (July 10, 2019), <https://www.nytimes.com/2019/07/10/opinion/facial-recognition-race.html>.

of equal opportunity in employment demands that the law should intervene. Given the limitations of traditional antidiscrimination laws to address the unlawfully discriminatory capabilities of automated hiring systems, a Lex Informatica derived framework, which provides a proactive ex ante approach of influencing design principles, would provide more meaningful governance of automated video interviewing systems.

REVISITING *ROOMMATES.COM*

G.S. Hans[†]

ABSTRACT

Fair Housing Council of San Fernando Valley v. Roommates.com holds an important place in the history of cases interpreting § 230, the federal law which has facilitated the growth of the modern internet. But unlike many other § 230 cases, which concern defamation claims, *Roommates.com* focused on alleged violations of the Fair Housing Act. The Fair Housing Act, the final major 1960s federal civil rights law, holds an important place in our racial justice history—one that § 230 and the *Roommates.com* decision limit.

This Article examines the history and legacy of *Roommates.com*, situating it within the framework of the Fair Housing Act to focus debates over § 230 reform. As a case that, at the time, complicated the dominant interpretations of § 230 and yet ultimately stymied enforcement of the Fair Housing Act online, *Roommates.com* demonstrates how the promise of civil rights laws has fallen short in a digital economy. Even as § 230 has facilitated speech online for individuals, civil rights protections have lagged. By re-evaluating *Roommates.com* in a larger history beyond technology law, this Article aims to evaluate more fully what § 230 reforms might further the goals of civil rights protections and what costs might result.

TABLE OF CONTENTS

I. INTRODUCTION	1228
II. THE SHARP TURN OF <i>ROOMMATES.COM</i>	1232
III. <i>ROOMMATES.COM</i> AND FAIR HOUSING ACT LITIGATION	1238
IV. PROMOTING THE FAIR HOUSING ACT IN A § 230 WORLD .	1243

DOI: <https://doi.org/10.15779/Z38DJ58H67>

© 2021 G.S. Hans.

† Associate Clinical Professor of Law, Cornell Law School. I am grateful to Cloe Anderson for her invaluable research assistance. Kurt Opsahl and Nicky Ozer patiently answered my obscure questions about the *Roommates.com* litigation. My colleagues Chris Serkin and Dan Sharfstein helped me understand the history and structure of the Fair Housing Act, providing excellent guidance to a novice. I thank Adam Cowing, Emma Llanoso, and Chris Morten for close readings and feedback, and Anupam Chander, Daphne Keller, Mason Kortz, Jef Pearlman, Jenn Prusak, and Blake Reid for their insights. I am grateful to the editors of the *Berkeley Technology Law Journal* for their thoughtful suggestions and assistance. All errors remain my own.

V. CONCLUSION1251

I. INTRODUCTION

Section 230¹ of the Communications Decency Act has transformed from a once-obscure federal statute to the *bête noire* of presidents past and present.² The law that facilitated the growth of the American digital economy has become the focus of critique from those of us—which is to say, basically everyone—dissatisfied with the current state of the internet.³ In addition to a spike in public attention, legislators and regulators have also turned their attention to § 230 in recent years, resulting in the enactment of the controversial SESTA/FOSTA legislation and the possibility of future legislative reforms.⁴

1. 47 U.S.C. § 230 (2018).

2. See Bobby Allyn, *As Trump Targets Twitter's Legal Shield, Experts Have a Warning*, NPR (May 30, 2020, 11:36 AM ET), <https://www.npr.org/2020/05/30/865813960/as-trump-targets-twiters-legal-shield-experts-have-a-warning> (“President Trump has a new rallying cry in his escalating crusade against Twitter. As he put it in a tweet Friday: ‘REVOKE 230!’ ”); The New York Times Editorial Board, *Joe Biden Says Age Is Just a Number*, N.Y. TIMES (Jan. 17, 2020), <https://www.nytimes.com/interactive/2020/01/17/opinion/joe-biden-nytimes-interview.html?smid=nytcore-ios-share> (“[The Times] can’t write something you know to be false and be exempt from being sued. But [Mark Zuckerberg] can. The idea that it’s a tech company is that Section 230 should be revoked, immediately should be revoked, number one. For Zuckerberg and other platforms.”)

3. See Todd Shields & Ben Brody, *Washington's Knives Are Out for Big Tech's Social Media Shield*, BLOOMBERG (Aug. 11, 2020, 1:00 AM PDT), <https://www.bloomberg.com/news/articles/2020-08-11/section-230-is-hated-by-both-democrats-and-republicans-for-different-reasons>.

4. SESTA/FOSTA, the Stop Enabling Sex Traffickers Act and Allow States and Victims to Fight Online Sex Trafficking Act, have been codified at 47 U.S.C. § 230(e)(5). See Aja Romano, *A new law intended to curb sex trafficking threatens the future of the internet as we know it*, VOX (July 2, 2018, 1:08 PM EDT), <https://www.vox.com/culture/2018/4/13/17172762/fosta-sesta-backpage-230-internet-freedom> (describing SESTA/FOSTA); *Wicker, Graham, Blackburn Introduce Bill to Modify Section 230 and Empower Consumers Online*, U.S. SENATE COMM. ON COM., SCI., & TRANSP. (Sept. 8, 2020), <https://www.commerce.senate.gov/2020/9/wicker-graham-blackburn-introduce-bill-to-modify-section-230-and-empower-consumers-online> (announcing legislation proposed by Republican Senators Roger Wicker, Lindsey Graham, and Marsha Blackburn to amend § 230); *Warner, Hirono, Klobuchar Announce the SAFE TECH Act to Reform Section 230*, MARK R. WARNER (Feb. 5, 2021), <https://www.warner.senate.gov/public/index.cfm/2021/2/warner-hirono-klobuchar-announce-the-safe-tech-act-to-reform-section-230> (announcing legislation proposed by Democratic Senators Amy Klobuchar, Mark Warner, and Mazie Hirono to amend § 230); *The Telecommunications Act's "Good Samaritan" Protection: Section 230*, DISCO, <https://www.project-disco.org/section-230/> (Oct. 21, 2021) (collecting proposals to amend § 230).

Enacted in 1996, § 230 provides immunity from liability for interactive computer services (often called platforms) that publish information from third parties.⁵ Such information (commonly referred to as user-generated content) might be potentially defamatory or otherwise unlawful, creating a risk of liability for the platforms for their role in publishing or hosting the unlawful user-generated content. At the scale at which the platforms hope to operate, the potential for liability would be immense, as would the costs of prescreening content.

But § 230 means that platforms cannot be held liable as the publisher or speaker of the content or for their screening provisions if those provisions are undertaken in good faith. Although § 230 has some exceptions—most notably, for federal criminal law and intellectual property law—it largely immunizes the platforms for their users’ content. Platforms can still be held liable for their own content.

Because of the relative brevity of § 230’s provisions—as Jeff Kosseff notes in his landmark study of the law, the core protections amount to merely twenty-six words⁶—caselaw interpreting the statutory language has played a central role in determining the scope of § 230’s protections. In an early case, *Zeran v. AOL*, the Fourth Circuit interpreted § 230 broadly, asserting that Congress had made a clear policy choice to limit platform liability.⁷ *Zeran* concerned a defamation claim that Kenneth Zeran, an AOL user, brought against the company for failing to remove postings at Zeran’s request. The postings falsely implied that he was selling tasteless t-shirts regarding the Oklahoma City bombing. AOL, though it had notice, did not promptly remove the postings. However, the Fourth Circuit found that § 230 still insulated the company from liability, as it was acting as a publisher.⁸

Soon after, other circuits followed the Fourth Circuit’s interpretation of § 230.⁹ However, the Supreme Court has never granted certiorari on a case

5. 47 U.S.C. § 230(c).

6. JEFF KOSSEFF, *THE TWENTY-SIX WORDS THAT CREATED THE INTERNET 2* (2019).

7. *See Zeran v. Am. Online, Inc.*, 129 F.3d 327, 330–31 (4th Cir. 1997).

8. *Id.* at 332 (“AOL falls squarely within this traditional definition of a publisher and, therefore, is clearly protected by § 230’s immunity.”).

9. *See, e.g., Ben Ezra, Weinstein, & Co. v. Am. Online Inc.*, 206 F.3d 980, 986 (10th Cir. 2000) (“[W]e agree with the Fourth Circuit’s decision in *Zeran*.”); *Batzel v. Smith*, 333 F.3d 1018, 1027–28 (9th Cir. 2003) (“Consistent with these provisions, courts construing § 230 have recognized as critical in applying the statute the concern that lawsuits could threaten the ‘freedom of speech in the new and burgeoning Internet medium.’” (citations omitted)); *Green v. Am. Online (AOL)*, 318 F.3d 465, 471 (3d Cir. 2003) (“By its terms, § 230 provides immunity to AOL as a publisher or speaker of information originating from another information content provider.”); *Universal Comm’n Sys., Inc. v. Lycos, Inc.*, 478 F.3d 413,

interpreting the law.¹⁰ As a result, courts have frequently ruled in favor of defendants (often, internet platforms), asserting that § 230 bars tort claims or statutory violations, relying upon *Zeran*'s broad reading of § 230 immunity.¹¹

Fair Housing Council of San Fernando Valley v. Roommates.com exists as a critical, partial exception to this trend.¹² *Roommates.com* is a rare appellate opinion to find that a website could potentially be held liable for developing and displaying unlawful user content despite § 230's broad immunization.¹³ *Roommates.com* involved a federal Fair Housing Act (FHA) claim (and associated California state law claims) brought against Roommates.com, a website designed to match potential renters with potential tenants. Given that Roommates.com facilitated connections between third parties, one might think that § 230 and *Zeran* would make this an easy case, as Roommates.com merely published third-party content rather than creating its own. But because Roommates.com mandated users to answer questions that could facilitate discrimination in violation of the Fair Housing Act and California state law, the U.S. Court of Appeals for the Ninth Circuit held in an en banc opinion that § 230 did not *completely* bar the plaintiff's claims.

Since this case created a rupture in the wall § 230 created and *Zeran* fortified, some plaintiffs have relied upon *Roommates.com* and its progeny to demonstrate why, despite § 230's broad grant of immunity, *their* claims should proceed.¹⁴ Few plaintiffs have been successful. Section 230 continues to stymie those hoping to find platforms liable for their choices or their users' content. For those who support the current regulatory regime governing platforms, *Roommates.com* is an outlier. If platforms do not get *too* involved in facilitating the users' content, § 230 remains effective. As a result, platforms generally avoid the pitfalls Roommates.com fell into.

418–19 (1st Cir. 2007) (“Although this court has not previously interpreted CDA Section 230, we do not write on a blank slate. The other courts that have addressed these issues have generally interpreted Section 230 immunity broadly. . . . In light of these policy concerns, we too find that Section 230 immunity should be broadly construed.”).

10. See *Biden v. Knight First Amendment Inst. at Columbia Univ*, 141 S. Ct. 1220 (2021) (Thomas, J., concurring), https://www.supremecourt.gov/opinions/20pdf/20-197_5ie6.pdf (citing critiques of § 230 on First Amendment grounds).

11. See, e.g., *Nemet Chevrolet v. Consumeraffairs.com*, 591 F.3d 250 (4th Cir. 2009) (finding that § 230 protected a website from defamation and statutory claims and upholding the trial court's dismissal of the plaintiff's complaint).

12. 521 F.3d 1157 (9th Cir. 2008).

13. *Id.* at 1175–76.

14. See, e.g., Complaint at 17, *Herrick v. Grindr, LLC*, 306 F. Supp. 3d 579 (S.D.N.Y. 2018), *aff'd*, 765 F. App'x 586 (2d Cir. 2019) (No. 17-CV-932) (arguing that “[t]he plain language of CDA § 230 does not extend to Grindr's operation and design”).

This Article complicates the common story of *Roommates.com* in technology law circles by discussing its relationship to the Fair Housing Act. Examinations of the case rarely describe how the underlying claim that *Roommates.com* sought to dismiss alleged violations of the Fair Housing Act—the last major civil rights law enacted in the 1960s and a cornerstone of Lyndon Johnson’s Great Society initiative. The civil rights laws of the 1960s, with their goal of promoting racial equality and securing equal rights for Black communities, represent a rare moment of federal legislative achievement to vindicate the rights of racial minorities. Yet this nexus between civil rights and technology in *Roommates.com* remains largely unexamined in technology law scholarship analyzing the case and its effects.

What would a world that expanded § 230’s carveouts for intellectual property and federal criminal laws to include specific civil rights laws (like the Fair Housing Act) look like?¹⁵ In considering these questions, this Essay aims to expand our understanding of how technology can work to improve rather than impede racial justice. How does *Roommates.com* appear when placed alongside the history of housing discrimination, the decades-long efforts by advocates and organizers to protect tenants, the unrealized promise of the Fair Housing Act, and the socioeconomic dynamics that have thwarted justice for the unhoused and under-resourced?

Long before § 230 came into being, the Fair Housing Act already had a large exception—the Mrs. Murphy exception, which exempts the Act premises that are owner-occupied and relatively small. However, this exception does *not* apply to housing advertisements. Effectively, under the Fair Housing Act, it is permissible for a landlord to discriminate against a potential tenant for a room in his house, so long as the landlord does not explicitly advertise in a discriminatory manner.

The Mrs. Murphy exception attempts to balance associational First Amendment rights against the Fair Housing Act’s purpose to promote housing equity and justice. But by prohibiting such small landlords from advertising in discriminatory ways, the Fair Housing Act also forestalls the *marketing* of a landlord’s potentially discriminatory preferences. In essence, while there might be associational or privacy interests for small landlords, there are also policy interests in preventing the spread of such preferences through advertising.

15. As a result of the carveouts, entities otherwise immunized for liability as publishers can still be held liable under federal criminal or intellectual property laws. 47 U.S.C. § 230(e) (setting forth exemptions). Thus, for example, the Department of Justice could file a case against a website that hosts content unlawful under federal law, like child pornography. *See* 47 U.S.C. § 230(e)(1).

Whether one agrees or not with that policy choice, it at least reflects some consideration of competing values.¹⁶ This lies in contrast with the blunt force application of § 230 and its near-total foreclosure of Fair Housing Act liability online, particularly for publishers. In essence, § 230 reflects an *unbalanced* approach to the tension between housing equity and speech online.

This Article proceeds in three parts. Part II discusses *Roommates.com* and how it complicated the trajectory of § 230's interpretation by courts. Part III situates *Roommates.com* among Fair Housing Act cases and larger debates regarding housing, technology, and race. Part IV concludes by considering avenues for changing the law to improve Fair Housing Act enforcement online and the strengths and weaknesses of each approach. Just as we should situate *Roommates.com* in a larger narrative of Fair Housing Act claims and civil rights litigation, reformers should also consider whether legislative reform of § 230 will address the massive problems the platforms present.

II. THE SHARP TURN OF *ROOMMATES.COM*

Section 230 was enacted in 1996 to avoid the repercussions of developing caselaw that might allow internet platforms to be held contributorily liable for the actions of their users.¹⁷ Although one major case prior to its enactment, *Cubby v. Compuserve*, held that platforms were not liable for potentially defamatory content posted by users, another—*Stratton Oakmont, Inc. v. Prodigy Services Co.*—imputed liability to the platform for user content.¹⁸

Section 230 creates a rule granting “interactive computer services” (ranging from sites that host user content to DNS providers that provide technical services) broad immunity from liability for the content of their users. It does this by expressly disclaiming secondary liability for platforms or users for the content of another, creating a safe harbor for content moderation decisions, and preempting state laws that run contrary to its protections.¹⁹ The first major test of § 230's protections, *Zeran v. America Online, Inc. (AOL)*, interpreted the scope of its language quite broadly.²⁰ *Zeran* concerned America Online's

16. See Norrinda Brown Hayat, *Accommodating Bias in the Sharing Economy*, 83 BROOK. L. REV. 613, 625 (2018) (describing the policy goals of the Mrs. Murphy exception); James D. Walsh, *Reaching Mrs. Murphy: A Call for Repeal of the Mrs. Murphy Exemption to the Fair Housing Act*, 34 HARV. C.R.-C.L. L. REV. 605, 605 (1999) (calling for abolition of the Mrs. Murphy exception).

17. KOSSEFF, *supra* note 6, at 2–3, 6.

18. *Stratton Oakmont, Inc. v. Prodigy Services Co.*, No. 31063/94, 1995 WL 323710 (N.Y. Sup. Ct. May 24, 1995).

19. 47 U.S.C. § 230 (c)–(e).

20. 129 F.3d 327 (4th Cir. 1997).

potential liability for allegedly defamatory content posted by an unknown third party on its service.²¹ AOL had notice of the content issue, but the Fourth Circuit held that notice did not impede AOL's ability to rely upon § 230's protections. The court observed that notice-based liability would create immense practical burdens for platforms, which it reasoned Congress had sought to avoid by enacting broad language in § 230 in the first place.²² Individual speakers would also suffer if notice-based liability were in effect, as a platform might, out of an abundance of caution, remove their content upon notice whether or not was defamatory, creating a chilling effect.²³

The *Zeran* decision is notable for its scope, as Mary Anne Franks, Jeff Kosseff, and others have observed.²⁴ It is very difficult for plaintiffs to file claims against platforms without running into § 230 and the caselaw following *Zeran's* broad interpretation. Of the cases that have not found § 230 to definitively preempt plaintiff claims, *Roommates.com* is among the most prominent.

Roommates.com began when two nonprofits opposed to housing discrimination filed suit against Roommates.com, a website that facilitated roommate matching. The nonprofits alleged that Roommates.com violated state and federal fair housing laws, among other statutes, by (1) allowing usernames that used words associated with race, religion, or sex; (2) allowing users to write essays describing what they were looking for in a roommate that could indicate discriminatory preferences; and (3) requiring a questionnaire that mandated disclosure of information regarding a user's age, gender, sexual orientation, occupation, and familial status.²⁵ The trial court focused primarily on the federal Fair Housing Act, which plaintiffs alleged Roommates.com had

21. *Id.* at 328.

22. *Id.* at 333.

23. *Id.* See, e.g., *Phila. Newspapers, Inc. v. Hepps*, 475 U.S. 767, 777 (1986) (discussing chilling effects as antithetical to First Amendment protections).

24. KOSSEFF, *supra* note 6, at 94–96; Mary Anne Franks, *How the Internet Unmakes Law*, 16 OHIO ST. TECH. L.J. 10, 19–20 (2020).

25. *Fair Hous. Council of San Fernando Valley v. Roommate.Com, LLC.*, 2004 WL 3799488, at *2 (C.D. Cal. Sept. 30, 2004), *rev'd and remanded sub nom.* *Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC.*, 489 F.3d 921 (9th Cir. 2007), *on reh'g en banc*, 521 F.3d 1157 (9th Cir. 2008), and *aff'd in part, vacated in part, rev'd in part sub nom.* *Fair Hous. Council of San Fernando Valley v. Roommates.Com, LLC.*, 521 F.3d 1157 (9th Cir. 2008). In the subsequent Ninth Circuit decision, the court noted that “[e]ach subscriber must also describe his preferences in roommates with respect to the same three criteria: sex, sexual orientation and whether they will bring children to the household.” *Fair Hous. Council of San Fernando Valley v. Roommates.Com, LLC.*, 521 F.3d 1157, 1161 (9th Cir. 2008). The site required its users to engage in selection based on these criteria. *Id.*

violated by allowing the publication of discriminatory statements, contravening section 804(c) of the FHA—a key section of the law applying to publishers. That provision makes it unlawful to:

make, print, or publish, or cause to be made, printed, or published any notice, statement, or advertisement, with respect to the sale or rental of a dwelling that indicates any preference, limitation, or discrimination based on race, color, religion, sex, handicap, familial status, or national origin, or an intention to make any such preference, limitation, or discrimination.²⁶

Plaintiffs effectively alleged that Roommates.com facilitated the unlawful content posted by its users as a condition of use, even if the website did not actually create or draft the content.²⁷ But the trial court applied Ninth Circuit precedent, which generally followed *Zeran*, to hold that § 230 precluded the FHA claim.²⁸ The court observed that though the plaintiffs were concerned that finding § 230 to preclude their claim might “eviscerate” the FHA, such a concern was not unique to the Fair Housing Act, given how broadly § 230 was written.²⁹

In other words, to the trial court, the case did not implicate the policy justifications underlying the FHA, given how broadly § 230 preempts liability (except for a few specific carveouts). Those carveouts, as the trial court noted, cover federal criminal law and intellectual property law and do not include any reference to civil rights or housing law.³⁰ Invoking one of the maxims of statutory construction, *expressio unius est exclusio alterius*, the court declined to read the FHA into that limited list.³¹

It is worth considering *why* federal civil rights laws are not part of the list of exceptions. Section 230 became part of a larger bill, the Communications Decency Act (CDA), which arose from concerns about explicit content online that minors might be able to access. Although § 230 was designed as an

26. 42 U.S.C. § 3604(c).

27. *Roommates.com*, 521 F.3d at 1166. (“By requiring subscribers to provide the information as a condition of accessing its service, and by providing a limited set of pre-populated answers, Roommate [sic] becomes much more than a passive transmitter of information provided by others; it becomes the developer, at least in part, of that information.”).

28. 2004 WL 3799488 at *3–5. Because the federal court only had jurisdiction over the state law claims if the federal claim survived, the court dismissed without prejudice the state law claims. *Id.* at *5.

29. *Id.* at *4.

30. *Id.* at *3.

31. *Id.*

alternative to the CDA—one that was supported by the American Civil Liberties Union (ACLU)—Congress ended up combining the two bills. Jeff Kosseff describes the drafters of § 230 as particularly focused on creating broad language that would attract minimal opposition from other constituencies, especially “law enforcement or industry groups,” while not creating so many exceptions as to poke holes in the law.³² Kosseff quotes one drafter:

We had to cut out criminal law. We had to cut out intellectual property. We knew the copyright people would kill us on that. It was a developing area. We were flying by the seats of our pants trying to make sure we got all the language in that we needed.³³

It is revealing that civil rights groups were not on the list of feared opponents. One view might be that because few people could really understand what § 230 would enable, fewer constituencies lobbied for changes to the proposed legislation. Kosseff describes the contemporaneous congressional milieu as relatively unaware of the dynamics or potential of the internet. Perhaps it is unsurprising that the content industries or the Justice Department *would* have understood what could happen if § 230 weakened their own legislative frameworks, given their awareness of the disruptive nature of new technologies.³⁴

The plaintiffs appealed the district court decision, and the Ninth Circuit partially reversed in a splintered ruling.³⁵ Judge Kozinski’s controlling opinion held that, because Roommates.com mandated a questionnaire that required the users to express impermissible preferences under the FHA, it could not wholly rely upon § 230 to immunize itself from liability.³⁶ In reconciling *Roommates.com* with earlier Ninth Circuit cases that had effectively applied *Zeran*, Kozinski noted that such cases did not necessarily grant § 230 immunity “to those who actively encourage, solicit and profit from the tortious and unlawful communications of others.”³⁷ This ruling created, as Kosseff notes, a potentially destabilizing result for online service providers on their home turf, as the Ninth Circuit covers both California and Washington State.³⁸

32. KOSSEFF, *supra* note 6, at 66.

33. *Id.*

34. *Id.* at 67–73. The lack of Congressional debate over § 230, as Kosseff reports, shows that few legislators really understood the ramifications of what they were voting on.

35. *Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC*, 489 F.3d 921 (9th Cir. 2007), *on reh’g en banc*, 521 F.3d 1157 (9th Cir. 2008).

36. *Id.* at 926–27.

37. *Id.* at 928 (discussing *Carafano v. MetroSplash.com*, an earlier Ninth Circuit § 230 case that relied upon the Ninth Circuit’s version of the *Zeran* framework).

38. KOSSEFF, *supra* note 6, at 174.

Roommates.com petitioned for, and obtained, an en banc rehearing. The ACLU of Northern California filed an amicus brief in support of neither party, seeking to balance the organization's commitments to both racial justice and to free speech, suggesting a path forward that the court ultimately took.³⁹ The ACLU argued:

Section 230 does not apply in this case to the extent that Roommate's liability is predicated upon the questions it asks in the questionnaire it designed, without respect to the responses provided by its members. Roommate is also not immune for its affirmative decision not to provide home seekers with email notifications of housing opportunities when the home seeker does not meet the allegedly discriminatory preferences of the provider. However, section 230 immunity does apply to members' statements in the comments sections of their profiles. That is content attributable solely to Roommate's members.⁴⁰

The ACLU's amicus attempted to balance the need for enforcing civil rights statutes like the FHA against the free speech concerns that could limit online expression. It casts the relevant inquiry as the platform's conduct (making active, mandatory choices in design to facilitate FHA violations, and thus acting like a speaker) against its users' choices (something largely out of the control of the platform).

Like the original three judge Ninth Circuit panel, the en banc court found that § 230 immunized some, but not all, of the claims that the plaintiffs had brought.⁴¹ In doing so, it did not wholly hew to the *Zeran* line of cases, but it also declined to drastically reform appellate interpretation of § 230: "The CDA does not grant immunity for inducing third parties to express illegal preferences. Roommate's own acts—posting the questionnaire and requiring answers to it—are entirely its doing and thus § 230 of the CDA does not apply to them. Roommate is entitled to no immunity."⁴² A website that does not passively host content, but actively induces it, could be liable.⁴³

39. See Brief of Amicus Curiae Am. C.L. Union of N. Cal. in Support of Neither Party, Fair Hous. Council of San Fernando Valley v. Roommates.Com, LLC, 521 F.3d 1157 (9th Cir. 2008) (Nos. 04-56916, 04-57173).

40. *Id.*

41. *Roommates.com*, 521 F.3d 1157 (9th Cir. 2008). As Judge Kozinski wrote both the controlling opinion in the initial panel and the en banc majority, it is unsurprising that the en banc opinion effectively ratified the earlier controlling opinion. The key language in the en banc opinion relies in large part on the initial panel's reasoning.

42. *Id.* at 1165.

43. Additionally, the site's search and email notification systems, which used the answers to its violative questions to filter and sort results, also fell outside of § 230's protections:

The en banc opinion opened the door to plaintiffs who earlier might have been absolutely foreclosed by *Zeran* and its progeny from overcoming § 230's grant of immunity. Now, if plaintiffs could allege that a platform had created content or induced third parties to violate the law, the platform could not rely upon § 230 as its Get Out of Litigation card. Though the en banc decision shifted the state of play for litigants, Judge Kozinski underplayed his holding: "The message to website operators is clear: If you don't encourage illegal content, or design your website to require users to input illegal content, you will be immune."⁴⁴

The decision did, in fact, change how plaintiffs brought cases, though its practical consequences were not as extensive as § 230 proponents might have feared. *Roommates.com* itself bounced around the district and appeals courts for a few more years, ultimately ending in a finding that the website had no liability as a publisher under the FHA.⁴⁵ Additionally, the selection of roommates did not violate the FHA.⁴⁶ Thus, no party (individual or website operator) was liable. The sputtering end to a high-profile saga serves as a symbol of the en banc opinion's effect upon § 230 caselaw—potentially revolutionary but in practice somewhat limited.

Kosseff notes that plaintiffs seeking to avoid § 230 will allege that *Roommates.com* governs, to "varying degrees of success."⁴⁷ Orly Lobel observes that platforms like Airbnb and Uber might be more comparable to *Roommates.com* than Craigslist, potentially exposing those platforms and their users to liability risks § 230 does not immunize.⁴⁸ And although a few cases have relied upon *Roommates.com* to preclude defendants from enjoying full

Similarly, Roommate is not entitled to CDA immunity for the operation of its search system, which filters listings, or of its email notification system, which directs emails to subscribers according to discriminatory criteria. Roommate designed its search system so it would steer users based on the preferences and personal characteristics that Roommate itself forces subscribers to disclose. If Roommate has no immunity for asking the discriminatory questions, as we concluded above, it can certainly have no immunity for using the answers to the unlawful questions to limit who has access to housing.

Id. at 1167 (citations omitted).

44. *Id.* at 1175.

45. 666 F.3d 1216 (9th Cir. 2012).

46. *Id.* at 1222 ("Because we find that the FHA doesn't apply to the sharing of living units, it follows that it's not unlawful to discriminate in selecting a roommate.")

47. KOSSEFF, *supra* note 6, at 180.

48. Orly Lobel, *The Law of the Platform*, 101 MINN. L. REV. 87, 145–46 (2016).

immunity under § 230, those cases are generally in the minority of § 230 cases.⁴⁹

III. *ROOMMATES.COM* AND FAIR HOUSING ACT LITIGATION

Although *Roommates.com* reshaped § 230 caselaw, it also has an important place in FHA litigation. Perhaps because of § 230's origins in responding to judicial decisions involving defamation claims, law and technology scholarship analyzing § 230 has tended to focus more on speech issues and less on civil rights claims like those in *Roommates.com*.⁵⁰ This Part contextualizes *Roommates.com* amongst Fair Housing Act cases—specifically, the provisions creating liability for publishing or printing allegedly discriminatory advertisements. This Part also considers the larger questions involving race, housing, and technology, both before and after the *Roommates.com* decision.

When it was decided, *Roommates.com* was not the only Fair Housing Act case involving § 230 in the federal courts. Less than a month prior to the en banc opinion in *Roommates.com*, the Seventh Circuit ruled in *Chicago Lawyers' Committee for Civil Rights Under Law, Inc. v. Craigslist, Inc.*⁵¹ The suit involved a similar challenge to *Roommates.com*—a nonprofit advocating for housing rights alleged that Craigslist was violating the FHA by allowing users to post notices that allegedly discriminated against potential renters.

Craigslist spends more time than *Roommates.com* discussing how, exactly, the Fair Housing Act might be violated if § 230 did not provide the website with immunity. Naturally, the most obvious FHA violation would occur if a landlord refused to rent to or treated inequitably a potential or actual tenant based on protected statuses like race, age, religion, or sex; the subsection of the FHA that prohibits such action is the first in a long list of prohibited

49. See *Huon v. Denton*, 841 F.3d 733 (7th Cir. 2016) (finding § 230 did not apply when website's employees allegedly created potentially defamatory content); *Lemmon v. Snap, Inc.*, 995 F.3d 1085 (9th Cir. 2021) (holding that § 230 did not preclude a negligent design claim); *Barnes v. Yahoo!, Inc.*, 570 F.3d 1096 (9th Cir. 2009) (finding that § 230 did not forestall plaintiff's breach of contract claim under a theory of promissory estoppel); *FTC v. AccuSearch, Inc.*, 570 F.3d 1187 (10th Cir. 2009) (holding that a website that acted to generate user content could not rely upon § 230); *Doe v. Internet Brands, Inc.*, 824 F.3d 846 (9th Cir. 2016) (finding that a failure-to-warn claim against a website operator was not precluded by § 230).

50. See, e.g., Eric Goldman, *Why Section 230 is Better than the First Amendment*, 95 NOTRE DAME L. REV. REFLECTION 33 (2019) (discussing the interaction between § 230 and the First Amendment in protecting online speech).

51. 519 F.3d 666 (7th Cir. 2008).

conduct.⁵² But liability also exists for *advertising* housing in a discriminatory manner under section 804(c) of the FHA.⁵³ Thus, a newspaper that allowed an advertisement for an apartment complex that only invited non-Black potential tenants to apply, could be subject to FHA liability independently of the apartment complex itself. It was under this theory that the nonprofits filed their claims against both Roommates.com and Craigslist.

Unlike Roommates.com, Craigslist won resoundingly at both the trial and appellate levels. The Seventh Circuit applied § 230's prohibition on publisher liability, determining that such language obviated any possibility for liability under section 804(c).⁵⁴ Although the Seventh Circuit was skeptical of *Zeran's* broad interpretation of § 230, it fully insulated Craigslist and did not foreshadow the Ninth Circuit's mixed decision for Roommates.com.⁵⁵

Historically, publisher liability cases under the FHA have involved advertisements or statements that a publisher made or that the landlord itself promulgated. In an early case, *U.S. v. Hunter*, the Fourth Circuit held that section 804(c)'s ban on discriminatory advertising applied to newspapers and survived First Amendment scrutiny.⁵⁶ A more recent case, *Iniestra v. Cliff Warren Investments, Inc.*, involved a landlord that printed notices and policies saying, in part, that children were required to stay in their apartments at certain times.⁵⁷ The plaintiffs successfully alleged that such a notice violated section 804(c)

52. 42 U.S.C. § 3604(a)–(b)

53. 42 U.S.C. § 3604(c).

54. 519 F.3d at 671 (“What § 230(c)(1) says is that an online information system must not ‘be treated as the publisher or speaker of any information provided by’ someone else. Yet only in a capacity as publisher could craigslist be liable under § 3604(c). It is not the author of the ads and could not be treated as the ‘speaker’ of the posters’ words, given § 230(c)(1).”).

55. *Id.* at 669 (“We have questioned whether § 230(c)(1) creates any form of ‘immunity.’”). Interestingly, Judge Easterbrook noted that the visibility of alleged housing discrimination on such sites might make FHA enforcement *easier* because civil rights organizations could monitor listings to see if any listings trigger FHA liability. *Id.* at 672 (“Using the remarkably candid postings on craigslist, the Lawyers’ Committee can identify many targets to investigate. It can dispatch testers and collect damages from any landlord or owner who engages in discrimination. It can assemble a list of names to send to the Attorney General for prosecution.” (citations omitted)). Judge Easterbrook does not note that, just as it would be difficult for platforms to prescreen potentially non-compliant postings at scale, it would likely be even more challenging for under-resourced civil rights organizations to monitor postings for potential FHA cases.

56. 459 F.2d. 205, 211 (4th Cir. 1972) (“[T]he congressional prohibition of discriminatory advertisements was intended to apply to newspapers as well as any other publishing medium.”).

57. 886 F.Supp. 2d. 1161, 1169–70 (C.D. Cal. 2012).

because it limited certain occupants' use of apartment facilities and thus discriminated based on familial status.⁵⁸

Such cases demonstrate clear prohibitions on publishing or printing materials that could fall within the FHA's restrictions on discrimination. Although a landlord of a sufficiently small property can rely upon the Mrs. Murphy exception to engage in potentially unlawful conduct in selecting tenants, a newspaper cannot print an advertisement from that landlord advertising such selection preferences because the exception does not cover section 804(c), as discussed below.

It is against this backdrop that we can see decisions like *Roommates.com* (in its ultimate disposition, in which the website was not liable under the FHA) and *Craigslist* as creating massive exceptions to the protection of the law. When aspiring renters moved from analog spaces (like newspaper classified ads) to online venues (like *Roommates.com*) in their housing search, they moved from a space governed by the FHA to one largely *ungoverned* by it. Although plaintiffs could pursue individual claims against posters, the platforms' practices that facilitate or encourage the underlying conduct are insulated from secondary liability. Potential renters likely did not even realize that shift. Yet, rather than considering *Roommates.com* as a destabilizing force to a major civil rights law, many technology law scholars focus on its effects on § 230.⁵⁹

Civil rights litigators, advocates, and scholars already grapple with exceptions to statutory protections. The Mrs. Murphy exception, which applies to both Title II (governing public accommodations) and the Fair Housing Act, effectively exempts from coverage premises that are owner-occupied and relatively small.⁶⁰ Such exceptions were designed to protect small establishments—those that run “small rooming houses all over the country,” in the words of one senator—from civil rights enforcement.⁶¹ They balance the associational rights of individuals in certain settings against the policy goals of civil rights protections. Crucially, the Mrs. Murphy exception does *not* apply

58. *Id.* at 1169.

59. See, e.g., Eric Goldman, *Roommates.com Denied 230 Immunity by Ninth Circuit En Banc (With My Comments)*, TECH. & MKTG. BLOG (Apr. 3, 2008), https://blog.ericgoldman.org/archives/2008/04/roommatescom_de_1.htm; Olivier Sylvain, *Intermediary Design Duties*, 50 CONN. L. REV. 203, 253–55 (2018) (discussing how *Roommates.com* has opened the door to plaintiffs seeking to hold platforms liable for their non-publishing decisions).

60. 42 U.S.C. § 2000a(b)(1) (“[O]ther than an establishment located within a building which contains not more than five rooms for rent or hire and which is actually occupied by the proprietor of such establishment as his residence; . . .”); 42 U.S.C. § 3603(b)(1).

61. See Norrinda Brown Hayat, *Accommodating Bias in the Sharing Economy*, 83 BROOK. L. REV. 613, 625 (2018); see also James D. Walsh, *Reaching Mrs. Murphy: A Call for Repeal of the Mrs. Murphy Exemption to the Fair Housing Act*, 34 HARV. C.R.-C.L. L. REV. 605, 605 (1999) (discussing the legislative origins of the Mrs. Murphy exception).

to section 804(c)—the section that governs discriminatory printed notices and advertisements, which *Roommates.com* and *Craigslist* implicate.

Roommates.com's parsing of liability for different components of *Roommates.com*'s site seems plausible given the different elements of § 230. The site had effectively no control over what users typed in its open field "Additional Comments," but it *did* influence user responses to its questions. Thus § 230 insulated *Roommates.com* from liability for the former but not the latter. For housing law experts, such parsing may seem ridiculous. Why should it matter to what degree the website enabled or induced potentially discriminatory content since it wouldn't matter for a traditional newspaper advertisement?

In the original Ninth Circuit panel opinion, Judge Reinhardt made a variation on this point in his concurrence:

On the final page of the sign-up process in which prospective users create their profiles, *Roommate's* site states, "We strongly recommend taking a moment to personalize your profile by writing a paragraph or two describing yourself and what you are looking for in a roommate" directly above a blank box. This page immediately follows the "My Roommate Preferences" form, which explicitly asks members to provide their preferences based on gender, sexual orientation and familial status. Judge Kozinski concludes that the "open-ended" recommendation on the "Additional Comments" page "suggests no particular information that is to be provided by members." Op. at 929. However, when viewed in the context of the entire sign-up process that conveys the message to prospective users that they should express their preferences for tenants based on race, gender, sexual orientation, national origin and religion, ordinary users would understand the recommendation to constitute a suggestion to expand upon the discriminatory preferences that they have already listed and to list their additional discriminatory preferences in that portion of the profile.⁶²

The opinions differ on how to characterize *Roommates.com*'s design choices. Judge Reinhardt argued that the "Additional Comments" section was more likely to give rise to potentially discriminatory content given how the site set up the field. In contrast, Judge Kozinski characterizes it as more "open-

62. *Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC*, 489 F.3d 921, 932 (9th Cir. 2007), *on reh'g en banc*, 521 F.3d 1157 (9th Cir. 2008), and *aff'd in part, vacated in part, rev'd in part sub nom.* *Fair Hous. Council of San Fernando Valley v. Roommates.Com, LLC*, 521 F.3d 1157 (9th Cir. 2008).

ended.”⁶³ This debate either seems completely vital to determining when and how § 230 applies or completely absurd—especially when contextualizing the debate against the FHA’s very broad prohibitions against publishing discriminatory advertisements, regardless of whether the publisher facilitated the development of the advertisement.

But recall that § 230, in its equally broad language, insulates information content providers from publisher liability. This creates the bizarre result in which a newspaper that reproduces a potentially discriminatory advertisement may be liable under the FHA for a print version, but not if it allows users to post advertisements on a website. In one sense, § 230 can be read as a “super-statute”—one that trumps almost all others.⁶⁴ As discussed later in Part IV, although this reading has dominated caselaw, following *Roommates.com*’s lead judges have expressed reservations about the broad immunity that *Zeran* established. In essence, treating § 230 as a super-statute seems less appealing now than it was at the time *Zeran* was decided.

Section 230 is an exception large enough to potentially swallow all of section 804(c). Like critics who have characterized the Mrs. Murphy exception as unnecessary, outdated, or discriminatory,⁶⁵ critics of § 230’s breadth highlight how it contributes to inequities.⁶⁶ It is also notable that the companies

63. Judge Reinhardt was on the en banc panel that reheard the case a year after writing this concurrence but joined Judge Kozinski’s majority en banc opinion without writing separately.

64. I use this term advisedly, referencing Justice Gorsuch’s description in his majority opinion in *Bostock v. Clayton County* of the Religious Freedom Restoration Act as a “super-statute.” See *Bostock v. Clayton Cty., GA*, 140 S. Ct. 1731, 1754 (2020) (“Because RFRA operates as a kind of super statute, displacing the normal operation of other federal laws, it might supersede Title VII’s commands in appropriate cases.”).

65. See Hayat, *supra* note 61, at 644 (“There are good reasons to amend Title II to remove the Mrs. Murphy exception, of course, including the fact that its very existence continues to signal that discrimination in some (even limited number) public accommodations is acceptable. The exception was rooted in racism and its modern-day proponents use it to perpetuate racism today.”); Walsh, *supra* note 61, at 607 (“The existence of an exemption for owner-occupied dwellings announces that our nation still tolerates discrimination. Implicit in the exemption is the belief that there is something so unsavory about Mrs. Murphy’s likely targets—African Americans, Latinos, Jews, families with children—that she should not have to live amongst them, even if they reside in separate units that she chose to make available on the market.”).

66. See, e.g., Danielle Keats Citron & Mary Anne Franks, *The Internet as a Speech Machine and Other Myths Confounding Section 230 Reform*, 2020 U. CHI. LEGAL F. 45, 54 (“Section 230 has subsidized platforms whose business is online abuse and the platforms who benefit from ignoring abuse. It is a classic ‘moral hazard,’ ensuring that tech companies never have to absorb the costs of their behavior. It takes away the leverage that victims might have had to get harmful content taken down.”); Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not*

that § 230 protects are mostly based in California, a state with the highest housing inequities in the nation.

Technology companies are hardly the sole cause of housing discrimination. The Fair Housing Act, on its own, cannot solve the housing inequities we see in America today. But § 230 immunity impedes the *goals* of the Fair Housing Act by essentially creating a largely nonregulable sector of the housing market. If the 1960s civil rights laws demonstrated a national commitment to racial justice and equality, § 230 makes that commitment incomplete. As a policy choice, § 230 is justifiable, given the need to provide broad protections to platforms to facilitate their development, growth, and moderation. The limited list of exceptions means that platforms can remain relatively confident that they are not exposing themselves to extensive litigation risk, theoretically incentivizing the creation of new platforms. But congressional choices mean that the values the Fair Housing Act was designed to protect and promote will never be fully realized under the current system—if they ever could have been at all.

IV. PROMOTING THE FAIR HOUSING ACT IN A § 230 WORLD

What, then, is to be done to address the problem of fair housing and discrimination in a country governed by § 230's protections? I envision three options: (1) changing how internet platforms advertise housing; (2) reinterpreting § 230's grant of immunity to such platforms qua their advertising activity; or (3) amending § 230. This Part evaluates each option in turn. Because it seems that legislative reform would best address the issues that *Roommates.com* highlighted, the preferred option is a statutory amendment that would include section 804(c) of the FHA alongside other exempted laws like federal criminal law and intellectual property law.

Given that § 230 has protected technology companies from a vast number of legal claims, private pressure—largely from individual users and civil society—has served as the main method for trying to urge companies to change their practices. Because such pressure is definitionally *ad hoc*, likely uncoordinated, and often from individuals or civil society (who generally have

Break: Denying Bad Samaritans § 230 Immunity, 86 FORDHAM L. REV. 401, 403 (2017) (“The CDA’s origins in the censorship of ‘offensive’ material and protections against abuse are inconsistent with outlandishly broad interpretations that have served to immunize platforms dedicated to abuse and others that deliberately host users’ illegal activities.”).

far less power than legislators or regulators) the likelihood of success varies widely.⁶⁷

In the housing context, the years since *Roommates.com* have given rise to a range of platforms that facilitate short-term uses, long-term rentals, and other housing arrangements—most notably VRBO, Airbnb, and Zillow. Though Airbnb primarily advertises itself as a platform facilitating short-term rentals, usually for vacationers, critics have inveighed against its effects upon local housing markets for years.⁶⁸ Those effects potentially include accelerating gentrification, pushing out long-time residents (who are often from historically marginalized groups), and limiting housing stock.⁶⁹

Airbnb has also received much criticism for the experiences of minority users, especially Black users, who have encountered racial discrimination while attempting to find accommodations. These experiences led to a prominent campaign, #AirbnbWhileBlack, which sought to force the company to improve its approach.⁷⁰ Because § 230 limits the possibilities for litigation activists and advocates must resort to advocacy efforts like #AirbnbWhileBlack to urge the company to change its practices that might, but for § 230, run afoul of the FHA. The company's policy responses have been voluntary, which—regardless of whether they are permanent, effective, or fully enacted—lack the financial pain and potential durability of a legal remedy.⁷¹

67. See Lee Rowland, *Naked Statute Reveals One Thing: Facebook Censorship Needs Better Appeals Process*, ACLU (Sept. 25, 2013, 10:07 AM), <https://www.aclu.org/blog/national-security/naked-statue-reveals-one-thing-facebook-censorship-needs-better-appeals> (discussing how the ACLU's influence and access facilitated its request to Facebook regarding a censorship decision).

68. See, e.g., Kyle Barron, Edward Kung & Davide Proserpio, *Research: When Airbnb Listings in a City Increase, So Do Rent Prices*, HARVARD BUS. REV. (Apr. 17, 2019), <https://hbr.org/2019/04/research-when-airbnb-listings-in-a-city-increase-so-do-rent-prices>.

69. See, e.g., Robert McCartney, *Airbnb Becomes Flash Point in the District's Hot Debate Over Gentrification*, WASH. POST. (Nov. 21, 2017), https://www.washingtonpost.com/local/airbnb-becomes-flash-point-in-the-districts-hot-debate-over-gentrification/2017/11/21/3c3bcd2-bf19-11e7-8444-a0d4f04b89eb_story.html.

70. See Hannah Jane Parkinson, *#AirBnBWWhileBlack hashtag highlights potential racial bias on rental app*, THE GUARDIAN (May 5, 2016, 10:28 EDT), <https://www.theguardian.com/technology/2016/may/05/airbnbwhileblack-hashtag-highlights-potential-racial-bias-rental-app> (discussing racial bias and disparate experiences for Black Airbnb users for both hosts and guests); see also Ray Fisman & Michael Luca, *Fixing Discrimination in Online Marketplaces*, HARVARD BUS. REV. (Dec. 2016), <https://hbr.org/2016/12/fixing-discrimination-in-online-marketplaces> (describing empirical research into race-based discrimination on Airbnb).

71. See *An Update on Airbnb's Work to Fight Discrimination*, AIRBNB (Sept. 10, 2019), <https://news.airbnb.com/an-update-on-airbnbs-work-to-fight-discrimination/> (discussing

Litigation addressing § 230's scope has had some success in tacking back from the expansive *Zeran* line. However, it is questionable that litigation could address the preemption of section 804(c) liability because existing cases don't address allegedly discriminatory advertising. Nevertheless, because recent cases show some willingness from courts to reorient the trajectory of § 230 caselaw, they signal the possibility of more meaningful legislative reforms.

In 2016, in a reversal of the typical § 230 case in which an aggrieved plaintiff sues a technology platform, only to run headlong into § 230, Airbnb proactively contested San Francisco's attempts to regulate Airbnb's rental procedures within the city. Airbnb argued that the possibility of criminal penalties against short-term rental platforms that failed to police their listings for compliance with city registration mandates ran contrary to § 230.⁷² Perhaps because platforms had been so successful in invalidating regulatory efforts under § 230, Airbnb initiated its case against the city in a pre-enforcement challenge.

Surprisingly, Airbnb lost its motion for a preliminary injunction. The district court held, relying upon *Roommates.com* and its Ninth Circuit progeny, that § 230 did not apply because the ordinance did not create speaker or publisher liability for platforms.⁷³ Specifically, because the ordinance required Airbnb to "monitor and police listings by third parties to verify registration" rather than create liability for Airbnb as the publisher or speaker for those listings, § 230 was not applicable.⁷⁴ In essence, the court asserted that the ordinance held "plaintiffs liable only for their own conduct—namely for providing and collecting a fee for Booking Services in connection with an unregistered unit."⁷⁵ Such decisions create the possibility that governments could regulate housing platforms without running afoul of § 230, although likely not to the extent that section 804(c) does, since that language explicitly

Airbnb's voluntary policy changes in response to public criticism). Because Airbnb's changes are self-regulatory, they lack the oversight and enforceability of a court order or consent decree.

72. *Airbnb, Inc. v. City & Cty. of San Francisco*, 217 F.Supp.3d 1066, 1071 (N.D. Cal, 2016). Although in practice and prior to beginning my academic work, I advocated publicly on behalf of Airbnb's position that § 230 immunized the company from liability under San Francisco's ordinance, that position was unsuccessful; Airbnb later settled with San Francisco. My advocacy was not funded by Airbnb, though we were in communication about the issue.

73. *See id.* at 1072–76.

74. *Id.* at 1072.

75. *Id.* at 1073. The court analogized the ordinance to a tax scheme that the City of Chicago levied on internet auction sites; that scheme was unsuccessfully challenged by StubHub! on § 230 grounds in the Seventh Circuit. *See City of Chicago, Ill. v. StubHub!, Inc.*, 624 F.3d 363, 366 (7th Cir. 2010).

creates publisher liability. Another way to frame this puzzle: the sorting, filtering, and other design actions of a platform aren't covered by § 230, but the more obvious forms of housing discrimination may be.

A later Ninth Circuit case, *HomeAway v. City of Santa Monica*, similarly upheld a local ordinance that required housing platforms to monitor their sites for listings that did not comply with the city's short-term rental registry.⁷⁶ As in *Airbnb*, the court held that § 230 did not preempt Santa Monica's ordinance because the ordinance did not require the platforms to monitor content.⁷⁷ The court relied upon *Roommates.com* and other Ninth Circuit decisions that limited § 230's scope in order avoid *Roommates.com*'s fears that expansive immunity would "create a lawless no-man's-land on the Internet."⁷⁸

Decisions like *Airbnb* and *HomeAway* demonstrate how courts have grown uneasy at the prospect of expansively reading § 230 to eliminate the ability of governments to regulate internet activities, particularly on critical issues like housing. This unease parallels judicial and scholarly concerns about rulings that use the First Amendment as a deregulatory tool.⁷⁹ Jeff Kosseff argues that judges have become skeptical of *Zeran* and how dominant the expansive view of § 230 became in the early years following § 230's enactment.⁸⁰ Judicial ambivalence—and extensive scholarly criticism—helps to explain why § 230 has received so much legislative attention in recent years.

Shortly after the *Airbnb* decision, Nancy Leong and Aaron Belzer argued that certain decisions by platforms (including encouraging users to post photos and creating rating systems) might implicate liability under civil rights laws,

76. 918 F.3d 676 (9th Cir. 2019).

77. *See id.* at 683 ("On its face, the Ordinance does not proscribe, mandate, or even discuss the content of the listings that the Platforms display on their websites."). Because the court construed the ordinance as not requiring the platforms to edit or monitor the content of listings but rather to check for licenses, it held that § 230 was not implicated.

78. 521 F.3d at 1164 (9th Cir. 2008).

79. *See, e.g.,* Amanda Shanor, *The New Lochner*, 2016 WISC. L. REV. 133, 135–38. Dissenting in a recent case, Justice Kagan observed the dangers of a deregulatory First Amendment:

Speech is everywhere—a part of every human activity (employment, health care, securities trading, you name it). For that reason, almost all economic and regulatory policy affects or touches speech. So the majority's road runs long. And at every stop are black-robed rulers overriding citizens' choices. The First Amendment was meant for better things. It was meant not to undermine but to protect democratic governance—including over the role of public-sector unions.

Janus v. Am. Fed'n of State, Cty., & Mun. Emps., Council 31, 138 S. Ct. 2448, 2502 (2018) (Kagan, J., dissenting).

80. *See* KOSSEFF, *supra* note 6, at 203–05.

including the FHA, without triggering § 230's protections.⁸¹ In effect, the ways in which platforms facilitate their users' content may remove their ability to rely upon § 230—at least if they are sufficiently active (such as with Roommates.com's dropdown menus). Leong and Belzer rely in part upon *Roommates.com*'s holding that when platforms materially develop content, § 230 immunity does not apply. Although this theory has not been tested in court, and Leong and Belzer do not address the final disposition of *Roommates.com* (in which the site was ultimately found not liable under the Fair Housing Act), it does provide an alternate method of promoting the goals of civil rights statutes by relying upon and extending § 230 and civil rights laws' existing interpretations without requiring statutory revisions.

Beyond the housing context, the Ninth Circuit analyzed more recently in *Lemmon v. Snap* a wrongful death claim against Snapchat, limiting the company's § 230 immunity.⁸² The plaintiffs alleged that Snapchat negligently designed a "Speed Filter" app that some users incorrectly perceived would reward them if they drove over 100 miles per hour. Tragically, two teenagers died in an automobile accident; their parents filed suit, claiming that the company knew about the myth regarding the Speed Filter and did not effectively limit the risk.⁸³ Snap claimed that § 230 immunized the company from liability.⁸⁴

The Ninth Circuit disagreed, applying *HomeAway* and an earlier case, *Barnes v. Yahoo*, to find that § 230 did not immunize Snap's design choices. In effect, the plaintiffs sought to hold Snap liable not for its actions as a publisher or speaker, but rather for the potentially negligent elements of its platform.⁸⁵ The Ninth Circuit ratified this approach, citing *Roommates.com* and noting that because the plaintiffs did not attempt to hold Snapchat liable for the content for another, but rather for its own choices, § 230 did not immunize the company.⁸⁶ In effect, the Speed Filter was analogous to the drop-down menus on Roommates.com.

81. Nancy Leong & Aaron Belzer, *The New Public Accommodations: Race Discrimination in the Platform Economy*, 105 GEO. L.J. 1271, 1306–99 (2017).

82. 995 F.3d 1085 (9th Cir. 2021).

83. *Id.* at 1087–90.

84. *Id.* at 1090.

85. *Id.* at 1092 ("It is thus apparent that the Parents' amended complaint does not seek to hold Snap liable for its conduct as a publisher or speaker. Their negligent design lawsuit treats Snap as a products manufacturer, accusing it of negligently designing a product (Snapchat) with a defect (the interplay between Snapchat's reward system and the Speed Filter). Thus, the duty that Snap allegedly violated 'springs from' its distinct capacity as a product designer." (quoting *Barnes v. Yahoo!, Inc.*, 570 F.3d 1096, 1107 (9th Cir. 2009), as amended (Sept. 28, 2009))).

86. *See id.* at 1093–94.

Though *Airbnb* and *HomeAway* demonstrate that platforms that implicate housing laws may not use necessarily § 230 as an escape hatch, and *Lemmon* shows that design choices are not immune under § 230, we are still a long way from enforcing the Fair Housing Act online, particular under section 804(c). Because section 804(c) explicitly invokes publisher liability, § 230 will inevitably immunize platforms that might otherwise be liable under that section, as in *Craigslist*, unless they engage in design conduct to a degree that arguably contravenes the FHA's protections. But since *Roommates.com* was ultimately found not liable even for its non-§ 230 protected conduct, the FHA itself needs more strength to effectively promote housing justice.

The broad protections of section 804(c) are likely irreplacable in an online environment unless Congress can both (A) craft an amendment that does not treat platforms as publishers (while still preserving some core § 230 protections) and (B) enact it. And although *Airbnb*, *HomeAway*, and *Lemmon* arguably widen the crack that *Roommates.com* created, courts generally continue to rely upon *Zeran* when assessing cases implicating § 230. Thus, advocates for robust civil rights law enforcement online have increasingly considered § 230 reform as the best path forward.⁸⁷

Many legislators have proposed various § 230 reforms. Because of the fast-moving nature of legislation, this Article considers more generally the *types* of reform proposals and how they might further effectuate the goals of the FHA (and perhaps other civil rights laws as well). I am sympathetic to the concerns that many have raised concerning § 230's role in creating an internet ecosystem rife with systemic injustice, horrifying conduct, and vast inequities. But my default has generally been to consider § 230 to be a bit like the old joke about democracy—the worst system, except for all the others. Although open to the possibility of reform, like others, I consider most of the recent proposals unworkable, counterproductive, or potentially unconstitutional.

For those, like me, who want robust civil rights laws and active enforcement of those laws, § 230 creates a vast realm where unenforceable conduct can run rampant. There are two potential modifications to § 230 that could further the goals of the FHA. Congress could change the statute in order to create a system in which design decisions—the ways in which an entity covered by § 230 sets up its system—are more explicitly unprotected (like

87. See, e.g., Ian Weiner, *Civil Rights Laws Will Significantly Benefit From Hirono, Warner, Klobuchar Section 230 Communications Decency Act Reform Proposals*, LAWS? COMM. FOR C.R. UNDER L. (Feb. 5, 2021), <https://www.lawyerscommittee.org/civil-rights-laws-will-significantly-benefit-from-hirono-warner-section-230-communications-decency-act-reform-proposals/> (expressing support for legislative proposals that would allow for greater civil rights enforcement online).

Roommates.com’s dropdown menus). Congress could also require prescreening for certain types of content (as in *Airbnb*, and akin to how newspapers must ensure print advertisements do not violate section 804(c)). Alternatively, the FHA, or perhaps only section 804(c), could be added to the list of statutes not covered by § 230, joining federal criminal and intellectual property law.

The design decisions approach has the benefit of hewing more closely to § 230’s policy goals, effectively making more explicit the interpretative moves of *Roommates.com* and *Lemmon*.⁸⁸ However, the design decisions approach fails to further the goals of the FHA in two key respects. First, section 804(c) would remain unenforceable against platforms that carried discriminatory advertisements unless the platform was careless enough to actively assist in the creation of those advertisements. A platform would need to be as foolish as *Roommates.com*, or as intransigent as *Airbnb* and *HomeAway*, to run afoul of the Fair Housing Act and other laws and ordinances that regulate housing. Most sophisticated companies would likely avoid such liability.

Second, to the degree that civil rights laws create a regulatory environment in which individuals can be relatively confident that they will not be discriminated against in public accommodations or housing, a design-oriented approach reverses the expectations. Rather than knowing *ex ante* that a platform likely opposes discriminatory housing policies, a user would need to litigate a case following a potentially discriminatory incident, shifting the burden. If civil rights laws are to have any force, they should create a regime in which individuals can feel confident that their rights are being protected.

Alternatively, legislators could add the FHA to the list of laws to which § 230 does not apply. This would make section 804(c) enforceable against platforms.

There are potential downsides to this approach. Most obviously, the SESTA-FOSTA debacle demonstrates how amending § 230, even for purported socially beneficial efforts, may create disastrous results for communities needing protection. The SESTA-FOSTA legislation supposedly addressed the problem of child sex trafficking, though in fact that problem was already addressed by existing federal criminal law (to which § 230 does not

88. Olivier Sylvain argues for this approach in a recent essay, drawing upon recent cases like *Lemmon* and *HomeAway*. See Olivier Sylvain, *Platform Realism, Informational Inequality, and Section 230 Reform*, 131 *YALE L.J. FORUM* 475, 501 (2021) (“My reform proposal is simple: online intermediaries should not be immune from liability to the extent that their service designs produce outcomes that conflict with hard-won but settled legal protections for consumers—including consumer-protection and civil-rights laws and regulations.”).

apply⁸⁹). As legislators considered SESTA and FOSTA, advocates and activists repeatedly sounded the alarm on how the legislation was both unnecessary, given existing law, and likely to cause harm to those who relied upon the community aspect of technology platforms to protect themselves.⁹⁰ Sex workers, who have relied upon online spaces to support their community, warned that the SESTA-FOSTA language could remove the possibility of sharing information and expressing solidarity online, driving sex work to the margins in which sex workers themselves would be more imperiled.⁹¹ Those concerns went unheeded, and the consequences, unfortunately, have come to pass.⁹² Beyond undoing the damage of SESTA-FOSTA, any future legislation modifying § 230 must do a far better job of avoiding negative consequences for those groups who lack the social or economic power to protect themselves.

The best way forward is to add section 804(c) to the list of exempted statutes, in addition to more meaningful reforms to the FHA itself. I leave proposals on the latter to more experienced housing scholars and advocates. As to the former, without allowing for FHA enforcement against discriminatory online housing advertisements—given how common such

89. 47 U.S.C. § 230(e)(1).

90. See Melissa Gira Grant, *Proposed Federal Trafficking Legislation Has Surprising Opponents: Advocates Who Work With Trafficking Victims*, THE APPEAL (Jan. 26, 2018), <https://theappeal.org/proposed-federal-trafficking-legislation-has-surprising-opponents-advocates-who-work-with-bf418c73d5b4/> (“Laura LeMoon, an anti-trafficking and sex workers’ rights advocate, wants to change that. She worries that legislation like SESTA and FOSTA, though ostensibly meant to help trafficking victims, is based on dangerous presumptions about the sex trade, which can actually harm both sex workers and people who are trafficked.”)

91. See *id.* (“Yet neither bill will result in justice for victims of human trafficking, anti-trafficking advocates and service providers told The Appeal. If passed, they say, the legislation stands to do more harm than good by failing to distinguish between trafficking victims and sex workers, eliminating sex workers’ source of income, and hampering anti-trafficking investigations.”).

92. See Melissa Gira Grant, *The Real Story of the Bipartisan Anti-Sex Trafficking Bill That Failed Miserably on Its Own Terms*, THE NEW REPUBLIC (June 23, 2021), <https://newrepublic.com/article/162823/sex-trafficking-sex-work-sesta-fosta> (“The new GAO report on SESTA/FOSTA, issued Monday, helps validate many of these concerns shared by sex workers and survivors of trafficking. As the report notes, rather than helping identify and prosecute traffickers, what SESTA/FOSTA did was push online sex work ads to the margins.”) Kendra Albert argues that for some lawmakers such consequences were the actual goal. See Kendra Albert, *Enough About FOSTA’s ‘Unintended Consequences’; They Were Always Intended*, TECHDIRT (July 29, 2021, 11:57 AM), <https://www.techdirt.com/articles/20210728/13245147264/enough-about-fostras-unintended-consequences-they-were-always-intended.shtml> (arguing that the legislation’s effects upon sex workers were intentional). I agree with this perspective.

advertisements are today—other adjustments will only allow for minor improvements.

A world in which platforms face section 804(c) liability would likely require a great deal of investment, both in terms of individual workers to prescreen content and automated review. Platforms might automatically ban all housing advertisements rather than prescreen. The housing market itself could change drastically. But the current dynamic, in which the already imperfect FHA lacks force online, effectively makes a powerful tool like section 804(c) into a dead letter.

I share the concerns of advocates who fear that § 230 reform will imperil the internet. Frankly, I am unsure whether such a legislative reform could help achieve the unrealized goals of the FHA. But the internet is already not functioning in so many obvious ways. Although this Article only discusses the FHA and related housing regulation, many other areas of traditional regulatory oversight remain largely unregulated online. That might be a situation we are willing to live with, or even embrace. But troubling consequences result from the current regime.

V. CONCLUSION

For technology lawyers and scholars, *Roommates.com* serves as an early indicator of how § 230 might not provide as much coverage to platforms as the early cases interpreting its text indicated. In revisiting it, my goal has been to employ a different lens to view its place not only in the history of § 230 cases but in a larger civil rights history as well. Because of the broad scope of § 230's language, it is easy to overlook the laws it overrides. Many § 230 cases deal with defamation, but the Fair Housing Act serves an important role in our civil rights history, one that § 230 lessens. Many scholars, advocates, companies, and governmental bodies are now taking a belated look at how technology and race intersect, while others who have done this work for years are finally receiving overdue attention.

It is impossible to say what will come of the calls and proposals for § 230 reform. But exempting section 804(c) would make a meaningful and expressive difference and could be worth the costs. No matter the outcome, we need a more sophisticated and expansive understanding of the role § 230 plays in a larger civil rights and racial justice movement. Only by clearly understanding current dynamics can we achieve real equity.

A TRIBUTE TO JOEL REIDENBERG

Paul M. Schwartz[†]

I. INTRODUCTION: IN THE BEGINNING

It began over Chinese food and continued as a friendship and a brotherhood for thirty years. Joel Reidenberg had just started teaching in 1990, when Marty Flaherty, a Fordham Law faculty member whom I knew, suggested that I give him a call. “We just hired someone in that same field as you,” Marty said with wonder. “What is it called again? Database protection?”

Marty also pointed out that while I had studied German law, Joel had studied French law, and we would probably have a lot to talk about. And we did have a lot to talk about—and for the next three decades!

Joel and I went out that day for Chinese food near Fordham, and we talked and talked, and a fast friendship was born. Joel was the most innovative and thoughtful of legal scholars, the most loyal of friends, and a sensitive and kind person. He was a Mensch.

II. A SCHOLAR AND POLICY ENTREPRENEUR

Regarding his scholarship, time and time again in his work, Joel clarified issues in law and technology in a way that set the terms of the research agenda for the rest of us. Indeed, the “Lex Informatica Symposium” of the *Berkeley Technology Law Journal* is a tribute to the lasting influence of one of his articles, the magnificent *Lex Informatica*.¹ In this seminal paper from 1997, Joel developed a series of insights about how information policy depends on network designs and systems architecture. But there is so much more to his scholarship than this one article.

In a different set of papers, Joel was skeptical from the start of internet commerce of industry self-regulation for privacy.² Here, he pointed to all the

DOI: <https://doi.org/10.15779/Z38B27PS1Q>

© 2021 Paul M. Schwartz.

[†] Jefferson E. Peysner Professor of Law, University of California, Berkeley, School of Law.

1. Joel R. Reidenberg, *Lex Informatica: The Formulation of Information Policy Rules through Technology*, 76 TEX. L. REV. 553 (1998).

2. Joel R. Reidenberg, *Setting Standards for Fair Information Practice in the U.S. Private Sector*, 80 IOWA L. REV. 497, 499–500 (1995) [hereinafter Reidenberg, *Setting Standards*]; Joel R. Reidenberg, *Privacy in the Information Economy: A Fortress or Frontier for Individual Rights?*, 44 FED. COMM. L.J. 195, 199 (1992).

ways that self-regulation of internet privacy would fail.³ Time has shown how correct Joel was in these articles.

Equally important are his writings about such diverse topics as student privacy, international trade, corporate privacy policies, and international data privacy law.⁴ Regarding student privacy, Joel ably drew on the resources of the Center on Law and Information Policy, which he had founded at Fordham Law School, to issue a stream of important papers. Among his critical work in this area was a co-authored empirical project role examining the commercial marketplace for student data and how privacy law failed to regulate it effectively.⁵ The important recommendations from this project received national media attention and led to Joel testifying on student privacy before Congress on three occasions as well as before state legislatures in Maryland, Oklahoma, and Virginia.⁶

Beyond student privacy, another important area of Joel's intellectual agenda concerned the connection between international trade law and privacy law.⁷ Many of Joel's articles were ones that my friend told me about as he was writing them, and, here, I must confess to a "Schwartz Delay." It was

3. He called the result of the self-regulatory approach the establishment of "a 'smoke screen' that in effect enables subtle, yet significant, manipulation of citizens through hidden control of private information." Reidenberg, *Setting Standards*, *supra* note 2, at 499–500. Joel also pointed to the weakness of privacy remedies. *See generally* Joel Reidenberg, *Privacy Wrongs in Search of Remedies*, 54 HASTINGS L.J. 877 (2003).

4. *See* N. Cameron Russell, Joel R. Reidenberg, Elizabeth Martin & Thomas Norton, *Transparency and the Marketplace for Student Data*, 22 VA. J.L. & TECH. 107 (2019) [hereinafter *Transparency and the Marketplace for Student Data*]; Joel R. Reidenberg, *Disagreeable Privacy Policies: Mismatches between Meaning and Users' Understanding*, 30 BERKELEY TECH. L.J. 39 (2015); Joel R. Reidenberg, *E-commerce and Trans-Atlantic Privacy*, 38 HOUS. L. REV. 717 (2001) [hereinafter *E-commerce and Trans-Atlantic Privacy*]; Joel R. Reidenberg, *Resolving Conflicting International Data Privacy Rules in Cyberspace*, 52 STAN. L. REV. 1315 (2000); Joel R. Reidenberg, *The Fundamental Role of Privacy and Confidence in the Network*, 30 WAKE FOREST L. REV. 105 (1995); Joel R. Reidenberg, *Rules of the Road for Global Electronic Highways: Merging the Trade and Technical Paradigms*, 6 HARV. J. L. & TECH. 287 (1993) [hereinafter *Rules of the Road*]; Joel R. Reidenberg, *Privacy in the Information Economy: A Fortress or Frontier for Individual Rights?*, 44 FED. COMM. L.J. 195 (1992).

5. *See generally* *Transparency and the Marketplace for Student Data*, *supra* note 4.

6. *See* *How Emerging Technology Affects Student Privacy: Hearing Before the Subcomm. on Early Childhood, Elementary and Secondary Educ. of the H. Comm. on Educ. and Workforce*, 114th Cong. (2015) (statement of Joel R. Reidenberg); *How Data Mining Threatens Student Privacy: Hearing Before the Subcomm. on Cybersecurity, Infrastructure Prot., and Sec. Tech. of the H. Comm. on Homeland Sec. and the Subcomm. on Early Childhood, Elementary and Secondary Educ. of the H. Comm. on Educ. and Workforce*, 113th Cong. (2014) (statement of Joel R. Reidenberg); *How Data Can be Used to Inform Education Outcomes: Hearing Before the H. Comm. on Educ. and Labor*, 111th Cong. (2010) (statement of Joel R. Reidenberg).

7. *See generally* *E-commerce and Trans-Atlantic Privacy*, *supra* note 4; *Rules of the Road*, *supra* note 4.

sometimes years after a paper was published, or a project was concluded, that I finally understood what Joel had been discussing with me. Trade law is a perfect example of this “Schwartz Delay”—and one that I had a chance to discuss with Joel in 2019 when I re-read his articles on this topic. The time has come to recognize how prescient these papers are. Back in the 1990s, Joel was already pointing at the disjunction between trade law and privacy law. This topic is now headline news with an important trade and privacy agreement between Japan and the European Union finalized in January 2019.⁸ This agreement has been followed by a similar combination, a trade agreement plus adequacy negotiation, between the European Union and South Korea.⁹ The adequacy component of this EU and South Korea discussion was only concluded on March 30, 2021.¹⁰ Finally, the law is trying to repair the “inherently unstable balance” between trade and privacy that Joel first identified in 1993.¹¹ Mea culpa, Joel, I now understand what you were talking about in those articles!

Joel was also a pioneer of comparative international privacy law. I was fortunate enough to be along for the ride and to work with him on two studies on behalf of the European Commission of the entity then called the European Community, and that now is the European Union. One study regarded the state of U.S. privacy law (not good),¹² and the second one concerned the already emerging lack of harmonization in EU data protection law.¹³

Our first project led to a 400-page-plus book, *Data Privacy Law* (1996), about U.S. information privacy law. Our conclusion? Regarding privacy law in the United States, we warned, “The narrow, dispersed approach to information regulation assumes that the treatment of personal information will be limited to one context within a particular industry of company. In reality, company information practices do not neatly fit within this sectoral thinking; there is widespread, cross-sectoral use of personal information.”¹⁴ This criticism of the sectoral approach to privacy law in the United States is still valid today. And here’s another quotation from this work, concerning our difficulty in gaining

8. *EU-Japan Trade Agreement Enters Into Force*, EUR. COMM’N (Jan. 31, 2019), https://ec.europa.eu/commission/presscorner/detail/en/IP_19_785.

9. *Joint Statement by Commissioner Reynders and Yoon Jong In, Chairperson of the Personal Information Protection Commission of the Republic of Korea*, EUR. COMM’N (Mar. 30, 2021), https://ec.europa.eu/commission/presscorner/detail/en/statement_21_1506.

10. *See id.*

11. *Rules of the Road*, *supra* note 4, at 290.

12. *See* PAUL M. SCHWARTZ & JOEL R. REIDENBERG, *DATA PRIVACY LAW: A STUDY OF UNITED STATES DATA PROTECTION* (1996).

13. *See* JOEL R. REIDENBERG & PAUL M. SCHWARTZ, *DATA PROTECTION LAW AND ON-LINE SERVICES: REGULATORY RESPONSES* (1998).

14. SCHWARTZ & REIDENBERG, *supra* note 12, at 379–80.

information regarding the specific practices of companies: “Companies are reluctant to risk embarrassment from public scrutiny of their practices.”¹⁵ Or, for a final example of how little things have changed on the privacy front since 1996: “In numerous instances, it is impossible for an average citizen in the United States to discover how, where, when, and why personal information is circulating.”¹⁶

The second study that Joel and I did for the Commission, *On-line services and data protection and privacy* (1998), demonstrates the emerging lack of harmonization under the Data Protection Directive of 1995 for the regulation of online services. In this second study, we analyzed four countries, which we divided among ourselves: Joel examined France and Belgium, and I looked at the United Kingdom and Germany. In our report, we called for European regulators to adopt a combination of “substantive data protection rules and principles with technical arrangements that allow the most efficient and least intrusive compliance.”¹⁷ Necessary as well was for European Union’s data protection officials to “have political input into the technical infrastructure decisions that affect the nature and characteristics of data flows.”¹⁸ Still sounds good twenty-four years later! Our assessment also found important differences in how these four Member States were responding to online services.¹⁹ This divergence provided evidence for a directly binding regulation from the European Union instead of its Data Protection Directive. The Union eventually followed this path with its General Data Protection Regulation of 2018, which, unlike a directive, has direct effect on the domestic law of Member States.²⁰

Joel’s comparative privacy law work continued over the next decades. One of his most important subjects was international data transfers. This subject is more timely than ever; international transfers of personal information have exponentially increased over the years and have been a continuing source of controversy. Already in 2001, Joel saw the shaky basis of the Safe Harbor Agreement between the Commission of the European Union and the United States. In Congressional testimony that year, he dismissed it as a “transitory

15. *Id.* at 389.

16. *Id.* at 390.

17. REIDENBERG & SCHWARTZ, *supra* note 13, at 149.

18. *Id.* at 153.

19. *Id.* at 121–37.

20. Regulation (EU) 2016/679 of the European Parliament and the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. L 119/1.

political success.”²¹ His analysis of the weakness of the Safe Harbor was confirmed in 2015 by the Court of Justice of the European Union, which in its famous *Schrems v. Data Protection Commissioner* (“*Schrems P*”) decision invalidated this trans-Atlantic agreement.²² As an international expert, Joel participated in two further studies for the Commission of the European Union.²³

Scholarship for Joel was not only a matter of paper, pen, and publication. My friend was a fighter and always stood up for what he believed in. He exemplified Congressman John Lewis’ concept of “good trouble.”²⁴ Regarding student privacy, for example, Joel addressed Congressional committees and state legislatures on this topic.²⁵ His work on financial privacy led him to testify before Congress in 2003, to serve as a Special Assistant Attorney General for the State of Washington from 2005–06, and to work as an expert consultant for the Federal Trade Commission in 1997. In 2004, he even testified in a hearing on the subject of credit reporting before the French National Privacy Commission (*Audition relative à la problématique des “centrales positives” en séance plénière du 13 mai 2004 de la Commission nationale d’informatique et des libertés*). When we worked on our reports for the European Commission, these projects took us to Washington, D.C.; Brussels, Belgium; Frankfurt, Germany; and Paris, France. It was an unforgettable experience to watch Joel navigating the corridors of power, including the West Wing of the White House, and to listen to him speak perfect French with officials of the European Union.

And while Joel was a diplomatic person, he knew when and how to modulate the sought effect. After he passed away, a mutual friend reached out to me to share his sorrow and grief. The friend was one with whom Joel and I had engaged in various privacy policy efforts, which involved occasional disagreements. He noted that Joel had a rare skill of managing to be

21. *The EU Data Protection Directive: Implications for the U.S. Privacy Debate: Hearing Before the Subcomm. on Commerce, Trade and Consumer Protection of the H. Comm. on Energy and Commerce*, 107th Cong. 72 (2001) (statement of Joel R. Reidenberg).

22. Case C-362/14, *Schrems v. Data Prot. Comm'r*, ECLI:EU:C:2015:650, P 98 (Oct. 6, 2015).

23. See J. SCOTT MARCUS, NEIL ROBINSON, LISA KLAUTZER, CHRIS MARSDEN, JOEL REIDENBERG, CAMILLA ABDER, CEDRIC BURTON, LISA COOMS, EZRA KOVER, YVES POULLET, FLORENCE DE VILLENFAGNE, FRANCK DUMORTIER, ADAM PEAKE, KEISUKE KAMIMURA & TAZUKO TANAKA, COMPARISON OF PRIVACY AND TRUST POLICIES IN THE AREA OF ELECTRONIC COMMUNICATIONS (2008); JAN DHONT, MARÍA VERÓNICA PÉREZ ASINARI, PROF. DR. YVES POULLET, JOEL R. REIDENBERG & LEE A. BYGRAVE, SAFE HARBOR DECISION IMPLEMENTATION STUDY (2004).

24. John Lewis (@repjohnlewis), Twitter (Jul. 16, 2019 8:44 AM), <https://mobile.twitter.com/repjohnlewis/status/1151155571757867011>.

25. See *supra* note 6 and accompanying text.

“simultaneously charming and annoying” and that he meant his comment “in the best kind of way.” I replied, “Well, if Joel was annoying, you deserved it!” The policy issues all came back to me, and I can only say that Joel was right to give our buddy a hard time about the issues in question.

While we are on the topic of “good trouble,” it is worth mentioning the Scalia-Reidenberg kerfuffle, which received widespread media coverage. Justice Antonin Scalia had objected when he learned that a Fordham law school class, using internet search engines, had created a lengthy dossier on his personal life. The matter began in January 2009, when Justice Scalia gave a speech in which he was quoted saying that “to treat much of the information on the web as private was ‘silly’ and that he did not care whether people knew what groceries he bought.”²⁶ At that moment, Joel’s “Information Privacy Law” course at Fordham was about to start a research exercise concerning the ways that technology can both invade and protect personal data. In a previous year, the exercise had sought to find “a specific piece of esoteric information” about Professor Reidenberg.²⁷ As Joel explained in a subsequent law review article, “During a class discussion early in the semester of Justice Scalia’s quotes about the silliness of privacy in his New York speech, the issues he raised about transparency made him a logical choice for the class research on a public figure.”²⁸

The resulting fifteen-page dossier gathered by students contained highly detailed information about Justice Scalia. As the *New York Times* observed, “the justice’s home address and home phone number, his wife’s personal e-mail address and the TV shows and food he prefers” were all in the class dossier, and all this information was collected from the internet.²⁹ In Joel’s summary:

This was precisely the teachable point Indeed to emphasize the value of the exercise as a pedagogical tool, the class dossier has remained a confidential, course document. None of Justice Scalia’s personal information was ever published or released by anyone in the class.³⁰

26. Joel R. Reidenberg, *The Transparent Citizen*, 47 LOY. UNIV. CHI. L.J. 437, 446 (2015) (quoting Jennifer Peltz, *Scalia Speaks on Digital Privacy at NYC Conference*, NEWSDAY (Jan. 28, 2009), <http://www.lawjournalbuffalo.com/news/article/current/2009/02/02/100308/scaliaspeaks-on-digital-era-privacy-at-nyc-conference>).

27. Reidenberg, *supra* note 26, at 447.

28. *Id.*

29. Noam Cohen, *Law Students Teach Scalia About Privacy and the Web*, N.Y. TIMES (May 17, 2009), <https://www.nytimes.com/2009/05/18/technology/internet/18link.html>.

30. Reidenberg, *supra* note 26, at 447.

The Justice was not amused, however, and told a website “what is legal may also be quite irresponsible.”³¹ Justice Scalia blasted Joel’s “judgment” in giving the class this exercise, which ignored that data brokers and others had already exercised their decision-making powers and their own judgment in making this information freely available.³² For Joel, the real point was, first, that the “governance dimension of the blending of publicly available and private information is poorly understood.”³³ The second lesson was the rise of the “transparent citizen” with its profound implications for democratic governance.³⁴

The issues that this exercise pointed to in 2009 are very much with us today. The Pentagon is now worried because cellphone database information for sale has revealed the presence of U.S. soldiers at a secret base in Syria.³⁵ Congress has a bill before it to protect the personal information of federal judges and those who share their residences.³⁶ Among its other provisions, the bill would prohibit commercial data collectors from selling, licensing, trading, purchasing, or providing judges’ personally identifiable information.³⁷ Joel would push back against these responses as too narrow and ask us to weigh the implications for all of the radical transparency caused by modern data handling practices.

During his life, Joel’s scholarship received numerous honors and its influence will continue to be felt for as long as people write about privacy and technology. Of his honors, I would first like to mention two because of their connection with Berkeley Law, my home institution. First, in 2013, Joel delivered the sixth annual privacy lecture at Berkeley Law. He spoke on “Data Access and Retention in the European Union and United States.”³⁸ Second, Joel received the Berkeley Center for Law & Technology’s Privacy Award in 2019. Due to the demands of his medical treatments, Joel was unable to receive the award in person, but his remarks on the big screen in International House demonstrated his sparkling wit and tremendous presence of mind. The award

31. Kashmir Hill, *Justice Scalia Responds to Fordham Privacy Invasion*, ABOVE THE LAW (Apr. 29, 2009), <https://abovethelaw.com/2009/04/justice-scalia-responds-to-fordham-privacy-invasion>.

32. *Id.*

33. Reidenberg, *supra* note 26, at 440.

34. *Id.* at 449–58.

35. Byron Tau, *Mobile-Phone Data Put U.S. Forces at Risk*, WALL ST. J., Apr. 27, 2021, at A8.

36. *See* S. 4711, 116th Cong. (2020).

37. *Id.* at § 4(c)(1).

38. The lecture later appeared in print. *See* Joel R. Reidenberg, *The Data Surveillance State in the United States and Europe*, 49 WAKE FOREST L. REV. 583 (2014).

was conferred for his “seminal scholarship, innovative policy entrepreneurship, and tireless support of the privacy community.”³⁹

The list of Joel’s many honors includes election as a Member of the American Law Institute, selection as an arbiter for the EU-U.S. Privacy Shield, and multiple awards for best privacy paper from the Future of Privacy Forum, a nonpartisan Washington, D.C. think tank. Joel also was a participating co-author on many computer science publications, all of which venture into technical realms into which most law professors would tremble to enter.⁴⁰ His speeches and addresses took him from Cambridge (MA) to Cambridge (UK) to Madrid to Bogoto to Amsterdam to Buenos Aires to Oslo to Montreal, and to multiple venues in his beloved Israel. His academic work is known and admired throughout the world, and his friends at universities and law schools are found on every continent.

III. A LOYAL FACULTY MEMBER AT FORDHAM AND A MENSCH

A picture of Joel Reidenberg would be incomplete without discussing Fordham University and his institutional-building efforts on its behalf. As a member of the Fordham Law faculty, Joel established its tech law center, the Center for Law and Information Policy. He also played a key role in founding the Samuelson-Glushko Intellectual Property Clinic at the law school.

If that were not enough, Joel took on an important institutional role at the university level. He served first as president of Fordham’s faculty senate and then as Associate Vice President for Academic Affairs, the latter a full-time position. In his role as an Associate Vice President for the University, he worked out of an office at the University’s Rose Hill campus in the Bronx. As a practical matter, these multiple roles meant an additional phone number to try when reaching out to Joel, who, despite his myriad roles, always found time for his friends.

39. 2019 BCLT PRIVACY AWARD, <https://www.law.berkeley.edu/research/bclt/bcltevents/2019bclt-privacy-lecture/2019-bclt-privacy-award/>.

40. See Sebastian Zimmeck, Peter Story, Daniel Smullen, Abhilasha Ravichander, Ziqi Wang, Joel Reidenberg, N. Cameron Russell & Norman Sadeh, *MAPS: Scaling Privacy Compliance Analysis to a Million Apps*, PROCEEDINGS ON PRIVACY ENHANCING TECH., July 2019, at 66; Alessandro Oltramari, Dhivya Piraviperumal, Florian Schaub, Shomir Wilson, Sushain Cherivirala, Thomas B. Norton, N. Cameron Russell, Peter Story, Joel Reidenberg & Norman Sadeh, *PrivOnto: A Semantic Framework for the Analysis of Privacy Policies*, 9 SEMANTIC WEB 185 (2018); Jaspreet Bhatia, Travis D. Breaux, Joel R. Reidenberg & Thomas B. Norton, *A Theory of Vagueness and Privacy Risk Perception*, 2016 IEEE 24th Int’l Requirements Eng’g Conference 26 (2016).

One tribute to Joel's loyalty to Fordham and to his boundless energy came from Bill Treanor, then the Fordham Law School Dean and now the Dean at the Georgetown Law Center. Dean Treanor once told me that, among his entire faculty, when he needed someone to attend an alumni event outside of the United States, the first call always was to Joel, the person who would make it happen. Joel would appear on the designated date, in the designated foreign country, at the designated venue—and with no more fuss than for a crosstown trip in Manhattan.

Another tribute to Joel as a member of the Fordham community came during a virtual memorial service for him in May 2020, which was held shortly after Joel passed away on April 21 of that year. Father Joseph McShane, the President of Fordham University, summed up my friend as “deeply embodying Jesuit values.” Here, one might be permitted a smile because Joel was so proud of being Jewish, which formed a central part of his identity. Intrigued by Father McShane's comment, however, I carried out research after the event regarding Jesuit values and found that if anyone fulfilled these virtues, it was my friend. Joel was truly a person who led “a virtuous life characterized by personal responsibility, respect, forgiveness, compassion, a habit of reflection, and the integration of body, mind, and soul.”⁴¹

Joel was also incredibly welcoming to other scholars. He was interested in ideas and debating their meaning and value. It did not matter if one was a Supreme Court Justice (see above) or a student at an international conference with whom he was talking over lunch. Joel was non-hierarchical, open to new concepts, and eager to invite newbies to the moveable feast that is information privacy law. I recall how at one meeting of the Privacy Law Scholars Conference, Joel pointed with amazement at a breakfast room at the Claremont Hotel in Berkeley where several hundred professors, government officials, and private sector lawyers were munching on muesli and muffins.

“Paul, remember when we were the only privacy law scholars in the country?” he asked with wonder. “Look at all these people.”

The increase in the ranks of our field delighted him, and Joel did everything to encourage young scholars and to help the field flourish. He was the best kind of senior scholar any academic field could have. His unspoken motto for privacy law was “the more, the merrier.”

Now I'd like to conclude with some final personal notes.

41. *Characteristics of a Catholic and Jesuit University*, FORDHAM UNIVERSITY, https://www.fordham.edu/info/20276/jesuit_and_catholic/647/characteristics_of_a_catholic_and_jesuit_university (last visited Jun. 2, 2021).

Joel and I always had a wonderful time together. Our work brought us around the world—and so many memories come back to me of presentations and meals and discussions. One important such meal-plus-policy effort took place at a restaurant on the Grande Place in Brussels. On that occasion, two important European privacy leaders joined us for mussels and Belgian beer on the night before a presentation of our research results to the Directorate-General responsible for data protection law. Joel and I also enjoyed fancy meals in France and pizza in New Jersey, and he always brought joy to the occasion.

And there was always technology to discuss—one favorite example was Joel showing me in the 1990's how he could use his telephone and a modem to email a digital document from one computer to another. Joel had brought his laptop into Fordham and sought to demonstrate how a digital file could go from his laptop over the internet to his office desktop, which was placed for this occasion directly next to his laptop. The idea was that he could send me drafts digitally in Fayetteville, Arkansas, where I was then teaching.

We sat in Joel's office. Dialing in, listening to the modem screech, and waiting for it to connect seemed to take forever. Plus, it was keeping us from our Chinese food. In no uncertain terms, I told Joel that I thought this process was a waste of our time and recommended that he simply mail me the next draft. My recommendation was as follows: "Just put the next version in an envelope, put a stamp on it; and I'll read it, write my comments on it, put the draft in a big envelope, and mail it back to you."

"We don't need the internet!" I sagely concluded.

Despite my implacable conviction at the time, it turns out that Joel was right about the future centrality of the internet—and so much more.

Joel was a tremendously brave person and a determined fighter as he faced illness. He was open about his health difficulties and challenges, and he shared news of his progress and setbacks in group-emails to his friends. His illness was nothing to be secretive about, but an aspect of life that he took on with optimism, determination, and humor. Joel was cosmopolitan, but the years in New Jersey were not wasted, and he was a Bruce Springsteen fan. One lyric that provided inspiration for him at this time was from the Boss: "No retreat, baby, no surrender." The other quotation of which he was fond came from Marshall Ferdinand Foch, who during the Battle of the Marne in 1914, informed his superior, "Pressed strongly on my right, my center gives way, impossible to move, excellent situation, I attack." Joel never retreated, he never surrendered, and, with the help of his physicians, he kept attacking.

When I think back on our relationship, another thing that comes to mind is how true a friend and wise an advisor Joel was. Joel called me "Big Bro,"

and I think he was avoiding calling me “Big Brother,” which, after all, would have been problematic for two privacy scholars. So, I was always “Big Bro,” and Joel was “brother” or “Brother Joel.” It was kind for Joel to give me this title, but the reality was that I depended on him more for advice and wisdom than he on me. Joel was the most generous and tactful of friends. He was my role model as husband, father, and, yes, brother. He was a loving and loyal husband to Pascale, and the proud father of Jeremy and David. The birth of his grandchildren Luca and Sophie gave him great happiness during the difficult period of his illness. In the photos that he sent me during this time of him with his grandchildren, his always warm smile seemed to reach from ear to ear. I do not know if I ever saw him happier.

I will miss Joel Reidenberg for the rest of my days, and I will continue to learn from his example as scholar, friend, and family man.

PRIVACY AS/AND CIVIL RIGHTS

Tiffany C. Li[†]

TABLE OF CONTENTS

I. INTRODUCTION	1265
II. BACKGROUND	1271
A. CIVIL RIGHTS IN U.S. LAW	1271
B. UNEQUAL ACCESS TO PRIVACY PROTECTION	1272
III. PRIVACY AS/AND CIVIL RIGHTS	1274
A. PRIVACY SUPPORTS CIVIL RIGHTS ADVANCEMENT.....	1275
B. PRIVACY AS CIVIL RIGHTS ACTIVISM	1276
C. CYBER CIVIL RIGHTS.....	1277
IV. A CIVIL RIGHTS FRAMEWORK FILLS GAPS IN PRIVACY LAW	1282
A. CONSTITUTIONAL PRIVACY AND EQUALITY PROTECTIONS	1282
B. SECTORAL PRIVACY LAWS	1286
C. THE PRIVACY TORTS	1290
V. RECOMMENDATIONS	1292
VI. CONCLUSION	1295

I. INTRODUCTION

When we talk about privacy only as a civil liberty, we erase those patterns of harm, that color of surveillance. And when we talk about privacy only as a civil liberty, we also ignore the benefits of privacy: Surveillance threatens vulnerable people fighting for equality. Privacy is what protects them and makes it possible.—Alvaro Bedoya¹

DOI: <https://doi.org/10.15779/Z381V5BF16>

© 2021 Tiffany C. Li.

[†] Assistant Professor of Law, University of New Hampshire Franklin Pierce School of Law; Affiliate Fellow, Yale Law School Information Society Project. The author thanks J. Sophia Baik, Alvaro Bedoya, Jill Bronfman, Jim Dempsey, Brittan Heller, Kristin Johnson, Thomas Kadri, Rebecca Lipman, Pedro Pavón, Rashida Richardson, Scott Skinner-Thompson for helpful feedback and commentary, Lucille Dai-He and Sophia Wallach for excellent editing, and the organizers and participants of the Privacy Law Scholars Conference and the *Berkeley Technology Law Journal* Symposium on Technology Law as a Vehicle for Anti-Racism.

1. Alvaro M. Bedoya, *Privacy as Civil Right*, 50 N.M. L. REV. 301, 306 (2020).

Surveillance is nothing new to black folks. It is the fact of antiblackness. —Simone Browne²

Decades have passed since the modern American civil rights movement began, but the fight for equality is far from over. Systemic racism, sexism, and discrimination against many marginalized groups are still rampant in our society. Tensions rose to a fever pitch in 2020, when a summer of Black Lives Matters protests, sparked by the police killing of George Floyd, an unarmed Black man,³ were met by a disturbing rise of white nationalist extremism⁴ leading to an attempted armed insurrection and attack on the U.S. Capitol on January 6, 2021.⁵ Asian-Americans faced rising rates of racism and hate crimes,⁶ spurred in part by inflammatory statements from the then-sitting President of the United States, Donald Trump.⁷ Members of the LGBT community faced attacks on their civil rights during the Trump administration,

2. SIMONE BROWNE, DARK MATTERS: ON THE SURVEILLANCE OF BLACKNESS 10 (2015).

3. See Larry Buchanan, Quoctrung Bui & Jugal K. Patel, *Black Lives Matter May Be the Largest Movement in U.S. History*, N.Y. TIMES (July 3, 2020), <https://www.nytimes.com/interactive/2020/07/03/us/george-floyd-protests-crowd-size.html>; Ray Sanchez, *Black Lives Matter protests across America continue nearly 2 months after George Floyd's death*, CNN (July 23, 2020), <https://www.cnn.com/2020/07/23/us/black-lives-matter-protests-continue/index.html>.

4. See Katanga Johnson & Jim Urquhart, *White nationalism upsurge in U.S. echoes historical pattern, say scholars*, REUTERS (Sept. 4, 2020), <https://www.reuters.com/article/us-global-race-usa-extremism-analysis/white-nationalism-upsurge-in-u-s-echoes-historical-pattern-say-scholars-idUSKBN25V2QH>.

5. See *Woman dies after shooting in U.S. Capitol; D.C. National Guard activated after mob breaches building*, WASH. POST (Jan. 7, 2021), <https://www.washingtonpost.com/dc-md-va/2021/01/06/dc-protests-trump-rally-live-updates/>.

6. See STOP AAPI HATE, STOP AAPI HATE: NEW DATA ON ANTI-ASIAN HATE INCIDENTS AGAINST ELDERLY AND TOTAL NATIONAL INCIDENTS IN 2020 1 (2021), https://secureservercdn.net/104.238.69.231/a1w.90d.myftpupload.com/wp-content/uploads/2021/02/Press-Statement-re_-Bay-Area-Elderly-Incidents-2.9.2021-1.pdf; CTR. FOR THE STUDY OF HATE & EXTREMISM, CAL. STATE UNIV. SAN BERNARDINO, REPORT TO THE NATION: ANTI-ASIAN PREJUDICE & HATE CRIME 3 (2021), <https://www.csusb.edu/sites/default/files/Report%20to%20the%20Nation%20-%20Anti-Asian%20Hate%202020%20Final%20Draft%20-%20As%20of%20Apr%2028%202021%2010%20AM%20corrected.pdf>.

7. See Kimmy Yam, *Trump can't claim 'Kung Flu' doesn't affect Asian Americans in this climate, experts say*, NBC NEWS (June 22, 2020), <https://www.nbcnews.com/news/asian-america/trump-can-t-claim-kung-flu-doesn-t-affect-asian-n1231812>.

including a rolling back of protections awarded to transgender individuals.⁸ There was also a sharp rise in antisemitism⁹ and Islamophobia.¹⁰

At the same time, the world faced a deadly pandemic that exposed the inequalities tearing the fabric of our society. Poor families struggled to survive the economic consequences of the Covid-19 coronavirus pandemic, and rural communities suffered from lack of access to healthcare.¹¹ Black and Brown communities in the United States suffered a disproportionate rate of sickness and death due to the coronavirus.¹² Women also faced disproportionate harms economically and otherwise, partially due to existing structural inequalities surrounding caretaking and careers.¹³ Individuals with disabilities often found themselves forgotten, as the push for expedient pandemic response overpowered even legal requirements for equal access.¹⁴ Internationally, the Global South also suffered disproportionately, perhaps most apparent in differences in access to medical supplies and vaccines.¹⁵

8. See Selena Simmons-Duffin, *'Whiplash' Of LGBTQ Protections And Rights, From Obama To Trump*, NPR: POLICY-ISH (Mar. 2, 2020, 3:12 PM), <https://www.npr.org/sections/healthshots/2020/03/02/804873211/whiplash-of-lgbtq-protections-and-rights-from-obama-to-trump>.

9. See *Antisemitic Incidents Hit All-Time High in 2019: ADL 2019 Audit of Antisemitic Incidents*, ANTI-DEFAMATION LEAGUE (May 12, 2020), <https://www.adl.org/news/press-releases/antisemitic-incidents-hit-all-time-high-in-2019> (“The American Jewish community experienced the highest level of antisemitic incidents last year since tracking began in 1979.”).

10. See Esther Yoon-Ji Kang, *Study Shows Islamophobia Is Growing In The U.S. Some Say It's Rising In Chicago, Too*, NPR (May 3, 2019), <https://www.npr.org/local/309/2019/05/03/720057760/study-shows-islamophobia-is-growing-in-the-u-s-some-say-it-s-rising-in-chicago-too>.

11. See Kelly A Hirko, Jean M Kerver, Sabrina Ford, Chelsea Szafranski, John Beckett, Chris Kitchen & Andrea L Wendling, *Telehealth in response to the COVID-19 pandemic: Implications for rural health disparities*, 27 J. AM. MED. INFORMATICS ASS'N 1816, 1816–18 (2020).

12. See Harmet Kaur, *The coronavirus pandemic is hitting black and brown Americans especially hard on all fronts*, CNN (May 8, 2020), <https://www.cnn.com/2020/05/08/us/coronavirus-pandemic-race-impact-trnd/index.html>.

13. See Courtney Connley, *More than 860,000 Women Dropped out of the Labor Force in September, According to New Report*, CNBC (Oct. 2, 2020, 2:45 PM), <https://www.cnbc.com/2020/10/02/865000-women-dropped-out-of-the-labor-force-in-september-2020.html>.

14. See Andrew Pulrang, *How Covid Relief Will Help Disabled People, and What Was Left out*, FORBES (Mar. 11, 2021, 2:08 PM), <https://www.forbes.com/sites/andrewpulang/2021/03/11/how-covid-relief-will-help-disabled-people-and-what-was-left-out/?sh=758512e314fc> (detailing the challenges for those with disabilities during the pandemic, such as students being unable to receive the assistance of one-on-one aides while schools provided only remote classes).

15. For references for the health disparities mentioned in this paragraph, the author discusses in greater depth the disparate impacts faced by marginalized communities in the pandemic in prior works, see Tiffany C. Li, *Post-Pandemic Privacy Law*, 70 AM. U. L. REV. 101,

The battle for civil rights is clearly not over, and the nation and the world have faced setbacks in the fight for equality. Most recently, the Supreme Court overturned *Roe v. Wade*, upsetting decades of national legal protection for the right to abortion, with implications for both privacy and equal protection.¹⁶ Meanwhile, the role of technology is also changing, with new technologies like facial recognition, artificial intelligence, and connected devices, offering new threats and perhaps new hope for civil rights.

To understand privacy at our current point in time, we must consider the role of privacy in civil rights—and, as scholars like Alvaro Bedoya have suggested, privacy itself as a civil right.¹⁷ This Article is an attempt to expand upon the work of privacy and civil rights scholars in conceptualizing privacy as a civil right and situating this concept within the broader field of privacy studies. This Article builds on the work of scholars like Anita L. Allen, Alvaro Bedoya, Khiara Bridges, Danielle Keats Citron, William Eskridge, Mary Anne Franks, Pamela Karlan, Scott Skinner-Thompson, and others who have analyzed critical dimensions of privacy and privacy law and advocated for changes in privacy law that can move our society forward to protect privacy and equality for all.¹⁸

To start, while much current scholarly discussion on privacy relates to privacy harms and negative rights against privacy violations, it is equally important to discuss privacy rights, including potential positive rights to privacy. If privacy can be a positive right awarded to individuals, what does it mean that not everyone is able to access that right equally? Perhaps it may mean that our conversation about privacy law must include ideas of equity, discrimination, and civil rights, in addition to the myriad other ways privacy rights are conceptualized currently. Under U.S. law, the right to privacy can be a “right to be let alone,”¹⁹ “right to autonomy,”²⁰ or vaguely, “the penumbra”²¹ of privacy rights. Privacy is a constitutional right, a civil liberty, a tort interest,

106–13 (2021); Tiffany C. Li, *Privacy in Pandemic: Law, Technology, and Public Health in the COVID-19 Crisis*, 52 LOY. U. CHI. L.J. 767 (2021) [hereinafter *Privacy in Pandemic*].

16. *Dobbs v. Jackson Women’s Health Org.*, 597 U.S. ____ (2022).

17. *See* Bedoya, *supra* note at 1 at 301.

18. This is a non-exhaustive list of inspirations and I remain grateful to the many scholars who have worked on these important, pressing issues of privacy and civil rights law.

19. Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARV L. REV. 193, 205 (1890); *see also* *Olmstead v. United States*, 277 U.S. 438, 572 (1928) (Brandeis, J., dissenting).

20. Daniel J. Solove, *Conceptualizing Privacy*, California Law Review (2002) at 1116 (explaining the connection between conceptions of rights to privacy and autonomy).

21. The penumbra theory of privacy is likely best encapsulated in *Griswold v. Connecticut*, 381 U.S. 479, 483 (1965) (discussing the “First Amendment [as having] a penumbra where privacy is protected from governmental intrusion”).

a consumer protection right, a contractual right, and, perhaps controversially, an intellectual property right.²²

However, as Scott Skinner-Thompson notes in his book, *Privacy at the Margins*, conceptualizing privacy “as a broad, amorphous, universalist value—something akin to autonomy, dignity, or personhood” leads to a “fail[ure] to capture the discrete, particular, and *material* harms that directly result from privacy violations.”²³ When privacy can mean almost anything, it can also mean almost nothing—and that is a problem, particularly for vulnerable populations who need privacy the most.

Existing privacy frameworks do not address the inequality in privacy rights afforded to different people.²⁴ For example, as a classical civil liberty, privacy protects the rights of assembly and speech for political dissidents and community advocates. However, conceptualizing privacy as only a civil liberty ignores the fact that these privacy rights, including the core rights to speech and assembly, have never been awarded equally to all Americans. If nothing else, a civil rights gloss on privacy rights can help us understand privacy in a way that might one day match protection for civil liberties.

Framing privacy primarily as a consumer protection right does not go nearly far enough to address the power imbalance between corporations and consumers²⁵ and does not address the ways in which consumer privacy rights and violations of those rights can shape laws and norms outside of the consumer context.²⁶ Furthermore, in the current data ecosystem, the public-private distinction can be unclear, and privacy harms can arise in grey areas not clearly governed by consumer protection law or any other specific area of law. For example, while laws like the Health Insurance Portability and Accountability Act (HIPAA) protect the privacy of health information,

22. For an analysis of the “privacy as intellectual property argument” (and why it arguably does not work), see generally Pamela Samuelson, *Privacy as Intellectual Property?*, 52 STAN. L. REV. 1125 (2000).

23. SCOTT SKINNER-THOMPSON, *PRIVACY AT THE MARGINS* 3 (2020).

24. For arguments that existing privacy frameworks necessitate further critique, see generally KHIARA M. BRIDGES, *THE POVERTY OF PRIVACY RIGHTS* (2017); VIRGINIA EUBANKS, *AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR* (2015); SKINNER-THOMPSON, *supra* note 23; Bedoya, *supra* note 1 at 301.

25. For more on the necessity of understanding larger structures of power and capital to understand the current status of privacy rights, see generally JULIE E. COHEN, *BETWEEN TRUTH AND POWER* (2019); SHOSHANA ZUBOFF, *THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER* (2018).

26. See Daniel J. Solove & Woodrow Hartzog, *The FTC and the New Common Law of Privacy*, 114 COLUM. L. REV. 583, 585 n.2 (2014) (analyzing consumer privacy law’s development and the role of the Federal Trade Commission, the federal agency leading the regulation of consumer privacy in the United States).

HIPAA only applies to particular covered entities (healthcare institutions, health plans, and healthcare clearinghouses), business associates, and particular types of information transmission.²⁷ HIPAA might not, for example, provide individuals with privacy protections for information they voluntarily share with a mobile phone application that provides counseling guidance, unless that application was linked to a covered entity or a covered entity's business associate.²⁸ In this circumstance, an individual might be interacting with the counseling application in a similar way as one would with a traditional mental healthcare provider. However, the patient might not receive the same level of privacy protection for the confidential information they share with the app. Arguably, HIPAA might not even apply to some health-related circumstances. Instead, lower standards of protection, like those in general consumer privacy law, but might be all that a healthcare patient is able to access.

Privacy rights are important for civil rights because the fight for civil rights is far from over. We have not nearly progressed enough in fighting discrimination in our government or in our society. Additionally, it is critical to consider privacy as a fundamental tool for protecting civil rights because our now technologically motivated world demands it. Technologies like facial recognition empower mass surveillance and discrimination at greater rates than ever before. Our increasingly connected world exposes all of us to ever more invasive violations of our privacy, both online and offline.

Technology in the wrong hands has great potential for harm. Even well-meaning technological pursuits may result in disparate harms to different marginalized groups. However, technology can also be a force for good. It is imperative that we consider the ethics of privacy and civil rights in developing new technologies and shaping the laws that regulate them to minimize the dangers and harms of technology and maximize the potential for technology to lead us to a more just and equal society.

This Article situates privacy in context with civil rights by exploring privacy as a civil right. Part I provides an overview of civil rights law in the United States and the unequal application of privacy protections to different people. Part II examines ways in which privacy can enhance or support civil rights and equality, including through the concept of cyber civil rights. Part III considers if, due to the inextricable links between privacy and civil rights protections, privacy itself ought to be considered a civil right. Part IV looks to framings of privacy in constitutional law, tort law, and sectoral consumer protection law and analyzes how a civil rights framework might resolve existing equity gaps

27. 45 C.F.R. § 160, 162, 164 (2013).

28. Health Insurance Portability and Accountability Act of 1996 (HIPAA), Pub. L. No. 104-91, 110 Stat. 1936 (1996).

in privacy law. Finally, Part V provides recommendations for how to better shape privacy law to protect both privacy and equality for all.

II. BACKGROUND

A. CIVIL RIGHTS IN U.S. LAW

Though the fight for equality is global (and must be global to truly reach its aims), this Article focuses primarily on American law. Under U.S. law, civil rights, broadly speaking, are rights granted by the government that protect citizens from discrimination by the government or other individuals.

One can generally understand “civil rights” as a set of rights that complements and contrasts with civil liberties, which are a set of individual freedoms enshrined in the Constitution that protect citizens against government intrusion on personal freedoms.²⁹ While the two terms are often oversimplified and conflated with one another, civil liberties grant you protections against government intrusion, and civil rights grant you protection against discrimination.

Much of our modern civil rights law is predicated on the Due Process Clause of the Fifth Amendment and the Due Process Clause and Equal Protection Clause of the Fourteenth Amendment. Influential laws codifying core civil rights protections include the Civil Rights Act of 1964, a landmark law that prohibits discrimination based on race, color, religion, sex, or national origin (“protected characteristics”), in public establishments that have a connection to interstate commerce or are supported by a state.³⁰ Of particular note are Title VII and Title IX of the Civil Rights Act. Title VII protects employees from employer discrimination based on the aforementioned protected characteristics, including age,³¹ pregnancy³² or maternity status,³³ sexual orientation,³⁴ transgender identity,³⁵ marriage or civil partnership status,

29. See, e.g., Christopher W. Schmidt, *The Civil Rights-Civil Liberties Divide*, 12 STAN. J. C.R. & C.L. 1, 1 (2016) (discussing the history of the civil rights and civil liberties divide in contemporary legal discourse).

30. Civil Rights Act of 1964, 42 U.S.C. § 1971 (1988).

31. Age Discrimination in Employment Act, 29 U.S.C. § 621 (1967).

32. Pregnancy Discrimination Act of 1978, 42 U.S.C. §§ 2000e.

33. *Phillips v. Martin Marietta Corp.*, 400 U.S. 542, 544 (1971).

34. *Bostock v. Clayton Cnty.*, 140 S. Ct. 1731, 1741 (2020).

35. *Id.* In *Bostock*, the Court specifically denotes “transgender status,” but one could reasonably argue that all gender identities should be covered under the but-for analysis used by the Court in this case. *Id.* at 1739–41.

or disability.³⁶ Title IX protects against discrimination on the basis of sex in schools and educational institutions that receive federal funding.³⁷

Other key civil rights laws include the Voting Rights Act (protecting against discrimination in access and ability to vote),³⁸ the Fair Housing Act (protecting against housing discrimination based on race, color, religion, sex, familial status, or national origin),³⁹ the Genetic Non-Discrimination Act (prohibiting discrimination on the basis of genetic information),⁴⁰ the Equal Pay Act (prohibiting wage discrimination on the basis of sex),⁴¹ as well as the Americans with Disabilities Act (protecting against discrimination for people with disabilities in areas including employment, transportation, and access to government services and public accommodations).⁴²

Outside of the federal context, there is, of course, a plethora of state civil rights laws as well. In addition, administrative agencies can enforce regulations that amount to civil rights protections, regardless of whether they are couched as civil rights laws. For example, the Federal Communications Commission regulates video accessibility standards, including obligations for television programmers to include closed captioning for disability access.⁴³ Effectively, these regulations can be understood to protect antidiscrimination rights of people with disabilities by mandating equal rights to access information.

Civil rights laws are useful because they recognize and attempt to prevent, mitigate, or correct harms suffered by different groups due to discrimination. Civil rights laws often focus on fundamental rights such as the right to vote and core social sectors such as employment and education. Privacy, a fundamental right, should be considered as integral for a person's civil rights protection as rights like employment and education. However, the relationship between privacy and civil rights has not always been clear, and privacy law often ignores crucial factors related to inequality and discrimination.

B. UNEQUAL ACCESS TO PRIVACY PROTECTION

This Article has already discussed the interplay between privacy as a civil liberty and privacy as a civil right. At first glance, privacy protections related to speech and assembly may seem like classical civil liberties, divorced from

36. Americans With Disabilities Act 1990, 42 U.S.C. §§ 12101–12213 (1990).

37. Title IX of the Education Amendments of 1972, 20 U.S.C. §§ 1681–1688.

38. Voting Rights Act of 1965, 52 U.S.C. § 10301.

39. Fair Housing Act, 42 U.S.C. § 3604.

40. Genetic Information Nondiscrimination Act of 2008, Pub. L. No. 110-233, 122.

41. Equal Pay Act, 29 U.S.C. § 206(d).

42. 42 U.S.C. §§ 12101–12213.

43. Twenty-First Century Communications and Video Accessibility Act of 2010, Pub. L. No. 111-260.

conceptions of civil rights. However, the United States has often awarded these privacy protections in unequal ways. Our constitutional privacy protections, the penumbra of privacy protections, do not adequately acknowledge the inequalities and inequities in privacy protections and privacy violations.

While surveillance can generate privacy violations and cause many people harm, these harms and violations often disproportionately injure people from marginalized populations. In her book *Dark Matters*, Simone Browne surveys the long history of surveillance against Black people in America, from the metaphor of the slave ship as a surveillance vessel to the phenomenon of TSA officers searching Black women's hair.⁴⁴ As Browne writes, "Surveillance is nothing new to black folks. It is the fact of antiblackness."⁴⁵

Other scholars have also researched the disproportionate impact of government surveillance on Black and Brown people, women,⁴⁶ the poor,⁴⁷ immigrants and undocumented people,⁴⁸ people with disabilities,⁴⁹ and more. In *The Poverty of Privacy Rights*, Khiara Bridges explains how poor mothers, especially Black and Brown mothers, lack privacy protections.⁵⁰ It is important to note that for every marginalized group that suffers from disproportionate privacy harms or civil rights violations, there are also individuals who belong to more than one marginalized identity. Kimberlé Williams Crenshaw's concept of intersectionality⁵¹ is key to understanding the discriminatory ways in which the law has awarded privacy protections and allowed for privacy violations. We must keep in mind the intersectional dimensions of civil rights violations, as well as privacy violations, if we are to work towards a theory of civil rights and privacy.

44. See generally BROWNE, *supra* note 2 (a comprehensive study of the impact of surveillance on Blackness and vice versa).

45. *Id.* at 10.

46. See generally ANITA ALLEN, *UNEASY ACCESS* (1988) (arguing for greater privacy protections for women).

47. See generally EUBANKS, *supra* note 24 (exploring the discriminatory impact of algorithmic systems on the poor).

48. *The Color of Surveillance: Government Monitoring of American Immigrants*, GEORGETOWN LAW SCH. CTR. ON PRIV. & TECH. (June 22, 2017), <https://www.law.georgetown.edu/news/web-stories/color-of-surveillance-immigrants.cfm>.

49. Elizabethette Guécamburu, *My Disability Doesn't Erase My Right to Privacy*, THE MIGHTY (Apr. 29, 2018), <https://themighty.com/2018/04/disability-and-the-right-to-privacy/>.

50. See generally BRIDGES, *supra* note 24.

51. See generally Kimberlé Crenshaw, *Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics*, 1989 U. CHI. LEGAL F. 139 (1989) (advocating for an intersectional analysis of race and gender to center the multidimensionality of Black women's experiences).

Unequal access to privacy is a civil rights problem. This is clearest when unequal privacy protections involve targeted surveillance of people from marginalized groups actively working to advance civil rights. For example, the United States has a long history of surveilling potential civil rights leaders, including civil rights legend Dr. Martin Luther King.⁵² In the 1960s, the FBI created COINTELPRO (an abbreviation of Counter Intelligence Program), a program aimed at surveilling Black civil rights leaders during the height of the civil rights movement.⁵³ While Americans today would likely like to consider themselves much more progressive than pre-civil rights movement Americans, it is clear that America as a country has not changed enough. In 2017, a whistleblower exposed another secret FBI program specifically aimed at surveilling Black activists as “Black Identity Extremists.”⁵⁴ Clearly, the battle for civil rights is not over, and privacy rights are still not afforded to all on an equal basis.

Modern civil rights law in the United States attempts to prohibit discrimination and equalize access to public accommodations and to rights we believe to be fundamental. Civil rights laws are an attempt to expose an inequity and remedy it. Today, our society faces a crisis of privacy discrimination in which people are unable to equally access their privacy rights, and some face disproportionate privacy violations and harms.

III. PRIVACY AS/AND CIVIL RIGHTS

Privacy is not only a foundational civil liberty, but also a core civil right. Conceiving of privacy as a civil liberty without understanding the interplay between privacy and equality does a disservice to both privacy and equality. As Alvaro Bedoya writes:

When we talk about privacy only as a civil liberty, we erase those patterns of harm, that color of surveillance. And when we talk about privacy only as a civil liberty, we also ignore the benefits of privacy:

52. See generally Hannah Giorgis, *When the FBI Spied on MLK*, ATLANTIC (Jan. 18, 2021), <https://www.theatlantic.com/culture/archive/2021/01/mlk-fbi-surveillance/617719/>; Benjamin Hedin, *The FBI's Surveillance of Martin Luther King, Jr. Was Relentless. But Its Findings Paint a Fuller Picture for Historians*, TIME (Jan. 18, 2021, 8:23 AM), <https://time.com/5930571/martin-luther-king-jr-fbi/>.

53. Julia Craven, *Surveillance of Black Lives Matter Movement Recalls COINTELPRO*, HUFFPOST (Aug. 20, 2015), https://www.huffpost.com/entry/surveillance-black-lives-matter-cointelpro_n_55d49dc6e4b055a6dab24008.

54. Sana Sekkarie, *The FBI Has a Racism Problem and It Hurts Our National Security*, GEORGETOWN SEC. STUDS. REV. (Aug. 19, 2020), <https://georgetownsecuritystudiesreview.org/2020/08/19/the-fbi-has-a-racism-problem-and-it-hurts-our-national-security/>.

Surveillance threatens vulnerable people fighting for equality. Privacy is what protects them and makes it possible.⁵⁵

When we talk about privacy only as a civil liberty, we erase patterns of harm from privacy violations that amount to or exacerbate discrimination and disparate impacts on marginalized populations. For example, while surveillance can lead to privacy violations and related harms for many people, these harms are often worse for marginalized populations. Privacy conceived as a civil liberty ignores the problems of unequal access to privacy and ignores the necessary place privacy has in creating the conditions for the fight for civil rights to continue.

A. PRIVACY SUPPORTS CIVIL RIGHTS ADVANCEMENT

Privacy is necessary for the fight for civil rights to continue. Privacy allows for the free association and assembly needed to organize and advocate for civil rights. Core to our understanding of democratic culture is the necessity of private spaces for gathering with others in pursuit of aims that may be political, especially those that may be politically unpopular.

Not only does privacy protect the political conduct of activists and civil rights workers, but privacy protects the personal and intellectual development necessary to spark the continued fight for civil rights. Privacy allows for the intellectual development of scholars, activists, and community leaders to create social change. Protecting privacy protects the intellectual freedom of individuals to explore new ideas and understandings.⁵⁶ Privacy allows individuals to interact with others in ways that generate democratic discourse and the conditions necessary for democratic participation.⁵⁷

Privacy protects the conditions necessary for a society to recognize, fight for, and protect civil rights. Earlier, this Article discussed the difference between civil liberties and civil rights.⁵⁸ Privacy supports fundamental freedoms like speech and assembly and the right against government intrusion into one's private affairs, and privacy is also critical in protecting civil rights. Privacy allows for the intellectual development, social interaction, and associational interactions necessary for societies to advance in civil rights.

Privacy protections are important for civil rights because the fight for civil rights is far from over. We have not nearly progressed enough in fighting

55. Bedoya, *supra* note 1, at 306.

56. See Neil M. Richards, *Intellectual Privacy*, 87 TEX. L. REV. 387 (2008) (explaining intellectual privacy).

57. See Julie E. Cohen, *What Privacy Is For*, 126 HARV. L. REV. 1904 (2013) (arguing that privacy can protect the ability to engage in democratic discourse).

58. See Schmidt, *supra* note 29.

discrimination, in our government or in our society. We must protect the privacy rights that foster the conditions through which individuals and groups can advance the cause for civil rights.

B. PRIVACY AS CIVIL RIGHTS ACTIVISM

Privacy, or the performance of privacy, can be an act of civil rights activism. In *Privacy at the Margins*, Scott Skinner-Thompson writes on the usefulness of privacy as expression of anti-subordination and political views.⁵⁹ He argues that marginalized populations, in particular, often utilize privacy as a form of protest against surveillance and discrimination.⁶⁰ Skinner-Thompson extends Judith Butler's theory of performativity⁶¹ to explain "performative privacy" as expressive conduct that ought to be understood as protected First Amendment expression.⁶² He gives examples including individuals using encryption technology to hide communications as a protest against government internet monitoring, as well as individuals wearing hoodies as protest against visual surveillance monitoring.⁶³

We can extend Skinner-Thompson's arguments even further and consider what a positive right to privacy as expressive conduct might be, particularly in light of the unique interactions between privacy and equality. Not only should individuals have the right to perform privacy, but they ought to be able to claim equal rights to the performance of privacy as a positive right. If there are laws or practices that make it difficult for some individuals to be able to perform privacy, then we should consider that to be a violation of First Amendment speech rights, privacy rights, and also equal protection and civil rights.

We can look to the example of the hoodie. Wearing a hoodie (a hooded sweatshirt or jacket) can be a form of protest—an act of performative privacy as civil rights activism. In 2012, 17-year-old Trayvon Martin, a Black boy, was shot dead by a neighborhood watchman while out buying Skittles. He was wearing a hoodie at the time.⁶⁴ Some stigmatized the hoodie as an article of clothing that either implied criminal association or could lead the wearer to be

59. See generally SKINNER-THOMPSON, *supra* note 23.

60. See *id.* at 45–107 (discussing performative privacy in theory and effect).

61. Skinner-Thompson discusses his inspirations in greater depth in Scott Skinner-Thompson, *Performative Privacy*, 50 U.C. DAVIS L. REV. 1673, 1690 (2017).

62. See SKINNER-THOMPSON, *supra* note 23, at 45–107 (discussing performative privacy in theory and effect).

63. Skinner-Thompson, *supra* note 61, at 1676.

64. CNN Wire Staff, *Police: Trayvon Martin's Death 'Ultimately Avoidable'*, CNN (May 18, 2012, 11:13 AM), <http://www.cnn.com/2012/05/17/justice/florida-teen-shooting>.

misidentified as such.⁶⁵ In the years after the fatal shooting, the symbolic meaning of a simple hoodie took on greater valence in American culture.⁶⁶ Because Martin had been wearing a hoodie when he was killed, a wave of protests included protestors who wore hoodies as a symbolic gesture, obscuring their identities, performing privacy and using the anonymity of the hoodie to show that what happened to Martin could happen again to anyone who looked like him or wore clothes like him, unless change occurred.⁶⁷

Here, we can understand the wearing of a hoodie to be an act of performative privacy as conceptualized by Skinner-Thompson—a protest against the mass surveillance and associated harms of discrimination in policing. However, if the act of wearing a hoodie comes with more risks (including being stopped by police) to some individuals based on the wearer’s race, then this form of performative privacy protest would not be equally accessible to all. A White woman may be able to wear a hoodie (and perform privacy) without fear of associated stigma, while a Black man may not be able to do the same.

If we understand performative privacy to be an act of civil rights activism, then this privacy right should be protected both as expressive speech and as civil right. Unequal access to privacy as a performative right then should be considered a civil rights problem that the law should remedy. Like housing, employment, and voting, privacy is a right that is fundamental to participation in our democratic society. And also like housing, employment, and voting, privacy is a right that is disproportionately awarded to and protected for some and not others.

C. CYBER CIVIL RIGHTS

The rise of the internet has changed many things, including our understanding of fundamental values like free speech, privacy, and equality. As more of life becomes intertwined with the internet and connected technologies, we must reconsider civil rights and how we can protect equality and antidiscrimination goals in online, offline, and hybrid contexts.

65. Amy Kuperinsky, *Hoodies: Danger or Fashion?*, NJ.COM (Mar. 30, 2019, 6:27 PM), https://www.nj.com/entertainment/2012/04/trayvon_martin_hoodie_march.html; MJ Lee, *Geraldo: Martin Killed Due to 'Hoodie'*, POLITICO (Mar. 23, 2012, 9:54 AM), <https://www.politico.com/story/2012/03/geraldo-martin-killed-due-to-hoodie-074392>.

66. Priya Elan, *Nine years after Trayvon Martin's killing, hoodies still spark debate*, GUARDIAN (Feb. 27, 2021, 3:00 PM), <https://www.theguardian.com/fashion/2021/feb/27/trayvon-martin-hoodies-black-young-people>.

67. Casey Glynn, *Trayvon Martin shooting sparks "hoodie" movement*, CBS NEWS (Apr. 2, 2012, 5:21 PM), <https://www.cbsnews.com/pictures/trayvon-martin-shooting-sparks-hoodie-movement/>.

In the classical sense, Warren and Brandeis conceptualized privacy as the “right to be let alone,”⁶⁸ particularly the right for wealthy elites like them to protect the private facts of their lives from the public eye.⁶⁹ This conception of privacy was created in a time when newspapers publishing society gossip pages were the primary threat to a person’s image. Warren and Brandeis did not conceive of deepfakes, nonconsensual sexual imagery, or the way that the internet would transform privacy violations related to defamation or harassment.

However, today, we live in a new world, where those harms exist and can impact a person’s access to civil rights. To address these harms, scholars like Danielle Keats Citron and Mary Anne Franks conceive of cyber civil rights as extending the doctrine of civil rights law from offline spaces to online spaces.⁷⁰ Citron and Franks focus some of their work on the critical problems related to harassment and other targeted harms related to the online speech environment.⁷¹ As they explain, technological privacy violations can impair a person’s ability to meaningfully participate in democratic society, and to exercise their rights as an individual.

Citron writes,

Civil rights laws are rarely invoked, even though cyber harassment and cyber stalking are fundamentally civil rights violations. Civil rights laws would redress and punish harms that traditional remedies do not: the denial of one’s equal right to pursue life’s important opportunities due to membership in a historically subordinated group.⁷²

Today, much of life is lived in the online space. Access to online spaces, then, is reaching a state of equal importance to access to offline spaces. When individuals do not have equal access to online spaces, due to a privacy violation that relates to a protected characteristic, we can and should understand this as a civil rights violation. For example, increased police surveillance in certain locations may make it uncomfortable for people from overly policed populations to spend time in those locations.

68. Warren & Brandeis, *supra* note 19, at 205.

69. For further discussion, see *infra* notes 117–118 and accompanying text.

70. See generally Danielle Keats Citron, *Cyber Civil Rights*, 89 B.U. L. REV. 61 (2009); Danielle Citron & Mary Anne Franks, *Cyber Civil Rights in the Time of COVID-19*, HARV. L. REV. BLOG (May 14, 2020), <https://blog.harvardlawreview.org/cyber-civil-rights-in-the-time-of-covid-19/>.

71. See generally Citron, *supra* note 70; Danielle Keats Citron, *Criminalizing Revenge Porn*, 49 WAKE FOREST L. REV. 345 (2014); Mary Anne Franks, *Sexual Harassment 2.0*, 71 MD. L. REV. 655 (2021).

72. DANIELLE CITRON, HATE CRIMES IN CYBERSPACE 23 (2014).

It is important to consider how we can protect civil rights in the online space and in the offline space. This is especially important when civil rights violations are connected to harms incurred online. For example, harassment against LGBTQ students in online spaces can lead to harassment in offline spaces, like public schools. It is empowering and novel to consider the extension of offline civil rights protections to online spaces. However, consideration is no longer enough. It is now time to take from the nascent doctrine of cyber civil rights to help solve technological privacy violations in the offline space as well.

It is now time to further extend the line of thinking within cyber civil rights into the offline realm as well. Today, the line between the cyber and physical realms is increasingly blurred. Moreover, extending the call for cyber civil rights to the equal protection of privacy in both online and offline spaces is a useful and cogent framework for the advancement of civil rights.

Civil rights laws create legal protections to prohibit discrimination against individuals. Some civil rights laws can afford individuals equal access to spaces and opportunities.⁷³ Cyber civil rights protections allow individuals equal access to online spaces and opportunities. The world of technologically driven civil rights violations exists neither purely online nor offline. One example of this is the phenomenon of mass surveillance. Mass surveillance is driven both by new technological dimensions of government surveillance as well as increasing consumer-led surveillance capitalism,⁷⁴ in which consumer-driven networked technology products create webs of surveillance.

If a person is barred entry into a public space on the basis of a protected characteristic, this could be considered a civil rights violation under modern civil rights law, including the Americans With Disabilities Act and the Civil Rights Act.⁷⁵ If a person is unable to access an online space due to policies or practices that amount to discrimination based on a protected characteristic, we can consider this a cyber civil rights violation. According to Citron and Franks, barriers to equal access do not have to be physical.⁷⁶ Indeed, policies and practices that allow for behavior with discriminatory effects can effectively bar a person from entry to an online space as a physical barrier in an offline space.

If a woman is refused entry into a public space due to her gender, it would be a civil rights violation—for example, a building with a sign saying, “Men

73. The Americans With Disabilities Act, for example, prohibits discrimination based on disability in access to public spaces. *See* 42 U.S.C. § 12101.

74. *See generally* ZUBOFF, *supra* note 25.

75. *See* Americans With Disabilities Act, 42 U.S.C. § 12101; Civil Rights Act of 1964, 42 U.S.C. § 1971. Both laws prohibit discrimination in public accommodations and public access.

76. *Id.*

only.” If a woman is unable to access an online space due to gendered harassment that makes the online space uncomfortable or dangerous, this could be a cyber civil rights violation—for example, a web forum rife with harassing posts targeting women. If a woman is unable to access an offline space due to technological privacy violations that make the offline space uncomfortable or dangerous, this, too, could be considered a civil rights violation—for example, a public space with large, visible cameras that record and stream to an open feed, with the promise that recorded images will be uploaded to the internet.

The same theoretical framework that could protect the expressive rights of individuals in an online space can also aid in protecting the privacy and speech rights of individuals in offline public spaces. In the past, one might conceive of civil rights violations concerning public spaces or public accommodations. One could also more easily separate the government and private companies as actors in civil rights disputes. However, today, with the advent of the internet and our increasingly connected world, humanity faces a new space—cyberspace. It can be difficult to determine the boundaries of any particular space online, raising questions about what constitutes a public space, a public forum, or a space for public accommodation.⁷⁷

Protecting privacy is necessary to protect civil rights, because much of our lives today are lived through technological means. This is particularly apparent now during the pandemic.⁷⁸ Today, much of society works online, studies online, and sometimes even celebrates, mourns, and worships online. To some extent, many people now spend much of their lives online. This increasingly connected world is causing a series of context collapses,⁷⁹ as the line between physical and virtual space is further eroded. While it is still necessary to protect privacy as appropriate for each informational and social interactional context,⁸⁰ it is also possible now to take the lessons from cyber civil rights and apply them to the offline world as well.

The more society moves to online spaces and the more individuals use connected technologies, the more opportunities there are for civil rights to be violated in non-traditional ways. Today, a growing body of jurisprudence concerns data-driven discrimination on social media websites. Researchers

77. *See, e.g.,* Robles v. Domino's Pizza, LLC, 913 F.3d 898, 905–06 (9th Cir. 2019) (analyzing whether ADA protections would require a restaurant's website to also be made accessible to people with disabilities).

78. *See Privacy in Pandemic, supra* note 15, at 804–06.

79. *Id.* at 784.

80. *See* HELEN NISSENBAUM, *PRIVACY IN CONTEXT: TECHNOLOGY, POLICY, AND THE INTEGRITY OF SOCIAL LIFE* (2009).

have found that Facebook and other similar advertising sites will often target only certain populations with ads for jobs or housing.

For example, in *National Fair Housing Alliance v. Facebook*, the National Fair Housing Alliance and other organizations sued social media platform Facebook, alleging that the platform allowed advertisers to exclude users of certain races from viewing housing ads.⁸¹ In the same year, the U.S. Department of Housing and Urban Development (HUD) filed a complaint against Facebook, similarly alleging that Facebook had practiced discrimination in regard to its housing ads.⁸² *NFHA v. Facebook* ended in settlement, without lasting precedent. In the meantime, journalists at The Markup have found that Facebook still allowed discriminatory advertisements on its platform as of 2020.⁸³ In the Facebook housing advertisement cases, the company's digital advertising practices arguably amounted to civil rights violations because they prevented individuals from protected classes from accessing housing opportunities.

Effectively, companies like Facebook can practice, enhance, or enable discrimination against marginalized and protected groups simply by hosting targeted ads. Targeted advertising generally relies on the collection and analysis of a bevy of data collected on or about an individual.⁸⁴ Thus, the discriminatory effect of job ads that exclude certain groups can, in one sense, be considered a privacy harm—i.e., a downstream harm that arises only due to invasive data collection. The discriminatory harms of targeted housing advertisements are inextricably linked to the privacy harms of bulk data collection and behavioral ad targeting. These are civil rights violations not simply due to the lack of access to housing opportunities but also the difference in strength and expansiveness of protection for the privacy of the users' data. In fact, one could argue that a Facebook users who had their data used to exclude them from accessing some housing ads effectively suffered a disproportionate privacy harm. This unequal privacy harm is not one that the law currently remedies.

More broadly, many of these joint online and offline civil rights violations can be linked to privacy violations, due to the interconnected nature of the

81. Complaint, Nat'l Fair Hous. All. v. Facebook, Inc., No. 1:18-cv-02689 (S.D.N.Y. Feb. 6, 2019).

82. See Charge of Discrimination, Dep't. of Hous. & Urb. Dev. v. Facebook, Inc., FHEO No. 01-18-0323-8 (filed Mar. 28, 2019).

83. Jeremy B. Merrill, *Does Facebook Still Sell Discriminatory Ads?*, THE MARKUP, (Aug. 25, 2020, 8:00 AM), <https://themarkup.org/ask-the-markup/2020/08/25/does-facebook-still-sell-discriminatory-ads>.

84. Rebecca Jennings, *Why targeted ads are the most brutal owns*, VOX, Sep. 25, 2018. <https://www.vox.com/the-goods/2018/9/25/17887796/facebook-ad-targeted-algorithm>

data ecosystem. Many online and offline civil rights violations now occur due to or are exacerbated by technological developments like big data, artificial intelligence, mass surveillance, facial recognition, and more. Digital privacy violations may harm all individuals whose data are collected or used. However, the misuse of individual data may disproportionately harm some people on the basis of protected characteristics. Protecting privacy, then, can protect individuals against these new technological (or cyber) civil rights violations.

IV. A CIVIL RIGHTS FRAMEWORK FILLS GAPS IN PRIVACY LAW

U.S. law has often treated privacy as a consumer protection right (in the sectoral privacy regime), as a civil liberty (in constitutional jurisprudence), or as a civil or criminal wrong (in [context]). Privacy has also been conceptualized as a property right and as a contractually determined right. These privacy frameworks all provide valuable ways to understand privacy in law and society, but they do not help us understand privacy protection as an antidiscriminatory or equalizing force. Framing privacy as a civil right is crucial because many privacy harms are disproportionately suffered by marginalized populations, and remedies are often inadequate.

A. CONSTITUTIONAL PRIVACY AND EQUALITY PROTECTIONS

Civil rights laws protect the vision of America as a country where “all men are created equal.”⁸⁵ The quest for civil rights has also been an influential factor in Supreme Court decisions on privacy, and it is possible to read civil rights into many major decisions that purport to focus on privacy. Privacy cases involving sexual conduct, contraception, and abortion necessarily involve civil rights, as the harms from restriction of these forms of conduct disproportionately harm individuals from certain protected classes, e.g., women and non-heterosexual people.

For example, *Lawrence v. Texas* invalidated criminal sodomy laws, partially based on a protected privacy interest in intimate sexual relations.⁸⁶ The majority opinion rejected arguments that the statute was invalid on Equal Protection Clause grounds. Instead, it focused on the due process violations apparent in a state law criminalizing a form of sexual conduct in a way that did not pass the strict scrutiny standard. Writing for the majority, Justice Kennedy argues:

85. In this phrase from the Declaration of Independence, as in elsewhere in the founding documents of our nation, we see the complex duality where structural inequalities in history and culture blinded or impeded a laudable quest for equality.

86. *Lawrence v. Texas*, 539 U.S. 558, 564–66 (2003).

The petitioners are entitled to respect for their private lives. The State cannot demean their existence or control their destiny by making their private sexual conduct a crime. . . . “It is a promise of the Constitution that there is a realm of personal liberty which the government may not enter.”⁸⁷

Arguably, this discussion of personal liberty as tied to respect for private lives is an extension of Supreme Court precedent on privacy, which continues to be vaguely and broadly constructed. Yet, the realm of personal liberty that the government may not enter is a clean construction of the civil liberty of privacy: the right against government intrusion into one’s private affairs.

Justice O’Connor’s concurrence in *Lawrence* takes another view.⁸⁸ O’Connor argues that the Texas law should be struck down not on due process grounds but due to its violation of the Equal Protection Clause. O’Connor argues that the Texas law is unconstitutional not because it improperly restricts personal liberty but because the law does not apply equally to all individuals. Instead, the restriction is tantamount to “a law branding one class of persons as criminal solely based on the State’s moral disapproval of that class and the conduct associated with that class.”⁸⁹ This contrasts with the majority opinion, which argues that the stigma of criminalization could remain even if the law was found unconstitutional for equal protection reasons.⁹⁰

One way to read *Lawrence* is to evaluate the majority opinion as a privacy opinion and the O’Connor concurrence as an equality opinion. However, this would be a limiting analysis of the case. Though the majority opinion focuses on liberty (and thus privacy), equality also appears to be a motivating factor. Indeed, Kennedy writes that “[e]quality of treatment and the due process right to demand respect for conduct protected by the substantive guarantee of liberty are linked in important respects, and a decision on the latter point advances both interests.”⁹¹ The Court makes clear the link between the right to privacy (the right to not have the government intrude upon your private conduct) and equality, where the privacy interest is linked to the conduct of a specific protected class.⁹²

Specifically, the Court says that any law criminalizing conduct specific to a protected class essentially discriminates against that class—“When homosexual conduct is made criminal by the law of the State, that declaration

87. *Id.* at 578.

88. *See id.* at 585 (O’Connor, J., concurring).

89. *Id.*

90. *Id.* at 575.

91. *Id.*

92. *See id.*

in and of itself is an invitation to subject homosexual persons to discrimination both in the public and in the private spheres.”⁹³ The majority opinion tells us that, regardless of equal treatment in the application of a law criminalizing a form of sexual conduct, the fact that the conduct in question was specific to “homosexual persons” means that criminalization of such conduct discriminates against those persons.

Privacy and equality are inextricably linked in many important cases. One can see echoes of the Court’s intentional melding of privacy and equality rights in cases like *Obergefell v. Hodges* that rely in part on *Lawrence*, as well as in cases both before and after *Lawrence* involving harms related to sex, contraception, and abortion.⁹⁴ For example, in *Eisenstadt v. Baird*, which the Court cites in *Lawrence*, the majority opinion relies in large part on an analysis of both the Due Process Clause and the Equal Protection Clause of the Fourteenth Amendment to protect the privacy rights of unmarried people using contraceptives.⁹⁵

Indeed, even in cases not specifically involving harms that are discriminatory toward a protected class, courts could potentially bring the values of civil rights and equality into discussions of privacy, e.g., the need to protect unpopular political groups in many First Amendment cases. For example, though this line of reasoning does not appear so much in the majority opinion, Sotomayor’s concurrence in *United States v. Jones* notes the effect of GPS surveillance as particularly harmful due to the potential for such sustained location surveillance in creating a record that “reflects a wealth of detail about [a person’s] familial, political, professional, religious, and sexual associations.”⁹⁶ Sotomayor cites *People v. Weaver*⁹⁷ to further explore the kinds of personal data such a record might entail:

93. *Id.*

94. *See generally* *Obergefell v. Hodges*, 574 U.S. 118 (2015) (holding that same-sex marriage was a right guaranteed under the Due Process and Equal Protection clauses of the Fourteenth Amendment); *see, e.g.*, *Griswold v. Connecticut*, 381 U.S. 479 (1965) (protecting the right of unmarried couples to purchase and use contraception); *Planned Parenthood v. Casey*, 505 U.S. 833 (1992) (upholding the constitutional right to abortion); *Eisenstadt v. Baird*, 405 U.S. 438 (1972); *United States v. Windsor*, 570 U.S. 744 (2013) (holding that a federal regulation banning same-sex marriage was unconstitutional under the Fifth Amendment’s Equal Protection clause); *Bostock v. Clayton Cty.*, 140 S. Ct. 1731 (2020) (holding that Title VII of the Civil Rights Act prohibited discrimination against employees based on sexual orientation).

95. *See Eisenstadt v. Baird*, 405 U.S. 438, 447–49 (citing *Griswold v. Connecticut*, 381 U.S. 479, 498 (1965) (White, J., concurring in judgment)) (1972).

96. *United States v. Jones*, 565 U.S. 400, 415 (2012) (Sotomayor, J., concurring).

97. *People v. Weaver*, 12 N.Y. 3d 433, 441–42 (2009).

Disclosed in [GPS] data . . . will be trips the indisputably private nature of which takes little imagination to conjure: trips to the psychiatrist, the plastic surgeon, the abortion clinic, the AIDS treatment center, the strip club, the criminal defense attorney, the by-the-hour motel, the union meeting, the mosque, synagogue or church, the gay bar and on and on.⁹⁸

What is interesting about the inclusion of this specific quote from the lower court opinion is that most of the examples it describes relate to conduct that is specific to protected classes. Trips to the psychiatrist may be more common for those with mental health disabilities. Plastic surgeons may see more patients who are women, and plastic surgery services are targeted more toward women.⁹⁹ Abortion clinics, AIDS treatment center, strip clubs, by-the-hour-motels, and gay bars similarly are places frequented by or targeted at people from protected classes due to sex, gender, sexual orientation, health status, and more. Mosques, synagogues, and churches are also frequented by specific groups due to and for the practice of religions, and religion is another protected characteristic. Union meetings are frequented by people with specific political aims, which are often protected under civil rights laws. Finally, due to systemic inequalities in policing and criminal justice, even trips to visit a criminal defense attorney may be more common for people from certain groups, particularly due to race.

One could extend this line of thinking to *Jones* and other cases by arguing that privacy and civil rights are inextricably linked. Keeping certain facts about oneself private may be more critical for people from protected classes. Note that civil rights laws protect certain “protected classes,” which is an imperfect analogue for protecting marginalized populations. Indeed, people from protected classes may require a higher threshold of privacy protections in order to achieve the same level of privacy and autonomy awarded to others. People who do not belong to marginalized groups may protect their own privacy with greater ease because the facts and behaviors they seek to keep private may be less prone to external interference or censure. It is far easier to use abortion clinic records to discriminate against women than against men, for example. Thus, some people may suffer from unequal access to privacy protection, just due to these people belonging to certain protected classes or even certain marginalized identities that do not fall under legally protected classification. If we consider privacy to be a public service, a public accommodation of a kind,

98. *Id.*

99. Sammy Sinno, Gretl Lam, Nicholas D. Brownstone & Douglas S. Steinbrech, *An Assessment of Gender Differences in Plastic Surgery Patient Education and Information in the United States: Are We Neglecting Our Male Patients?*, 36 AESTHETIC SURGERY J. 1 (2016).

or at the very least, a right the public ought to have, then privacy must be protected as a matter of civil rights and equality. The fact that types of personal information can identify someone as being a member of a protected class makes the privacy interest so demonstrably clear and necessary to protect.

It can also be useful to understand the thread of civil rights values across different strands of legal thought on privacy. When arguing that equal protection and privacy are inextricably linked in women's rights cases, Elizabeth M. Schneider suggests "concepts of equality are necessary for a robust understanding of privacy, and concepts of privacy are necessary for the full realization of equality."¹⁰⁰ Indeed, an understanding of privacy that lacks acknowledgement of inequity and discrimination cannot truly protect privacy for all.

B. SECTORAL PRIVACY LAWS

Outside of constitutional rights, privacy in the United States is often protected as a tort interest and as a consumer protection right. Privacy rights can also be protected in criminal law, e.g., criminal laws against stalking. While hopefully few individuals will ever find themselves fighting for or against tort or criminal privacy claims in a court of law, an increasing number of people may find themselves the main subject of another area of privacy law: privacy as a consumer right.

There is no federal privacy law in the United States, which differentiates the United States from the European Union, where the General Data Protection Regulation (GDPR)¹⁰¹ acts as an omnibus regulation that attempts to comprehensively govern data protection throughout the EU member states.¹⁰² In the absence of a U.S. federal privacy law, much of what constitutes U.S. privacy law becomes a matter of federal sectoral privacy laws, federal administrative codifications of privacy principles, and state privacy laws. Many of these state laws frame privacy as a consumer protection right.

With the modern omnipresence of technology, companies like Facebook, Google, and Apple—a smart phone in every pocket, a camera on every street corner—privacy has become a buzzword, so hotly debated and loudly degraded that it has, in some ways, become a concept almost devoid of

100. Elizabeth M. Schneider, *The Synergy of Equality and Privacy in Women's Rights*, 2002 U. CHI. LEGAL F. 137, 138 (2002).

101. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

102. See generally Meg Leta Jones & Margot E. Kaminski, *An American's Guide to the GDPR*, 98 DENVER L. REV. 1 (providing a succinct overview of the GDPR aimed at U.S. readers).

meaning.¹⁰³ Whether consumer, user, citizen, or human being, most of us harbor some implicit understandings of privacy, but it would be hard to argue that humanity has reached a general consensus on how best to protect privacy. This is particularly apparent in the United States, despite our nation's self-belief of ours as a nation above all others in devotion to freedom and liberty.

U.S. federal sectoral privacy laws regulate privacy for different industries, often focusing on harms to consumers. Some federal sources of authority for privacy as consumer protection law include the Fair Credit Reporting Act (FCRA),¹⁰⁴ the Health Information Portability and Accountability Act (HIPAA),¹⁰⁵ the Children's Online Privacy Protection Act (COPPA),¹⁰⁶ and general Federal Trade Commission (FTC) privacy common law.¹⁰⁷ Still, privacy laws that attempt to protect consumers as a monolith do not effectively protect all consumers, because the harms suffered by individuals are not suffered equally.

Furthermore, the sectoral nature of U.S. consumer privacy protection law provides unequal access to privacy protection for different classes of individuals and groups, as well as to different environments and against different actors. Some of the unequal access to privacy protections comes down to the fact that many consumer protection laws rely on individual consumers raising complaints to agencies or filing suits against companies. The legal process is inaccessible for many, and any form of privacy protection that relies on a legal system rife with inequality will likely also produce unequal results.

This consumer rights framework for privacy can also be insufficient because the growing collaboration between public and private actors can cause disparate harms and the way that connected technologies function make the larger data ecosystem more difficult to govern. For example, Amazon marketed its Ring camera products directly to consumers as home security systems.¹⁰⁸ However, the company also marketed to police departments. In

103. This is not to say that privacy is dead but merely that the myriad of unresolved questions about privacy, a myriad aided and abetted by changing social understandings of technology, contributes to a bit of confusion for the average consumer (or the average lawyer), that is, what privacy is and what is it for, and what is it not and what is it not for?

104. Fair Credit Reporting Act, 15 U.S.C. § 1681.

105. HIPAA, 45 C.F.R. §§ 160, 162, 164.

106. Children's Online Privacy Protection Act, 15 U.S.C. §§ 6501–06.

107. See Solove & Hartzog, *supra* note 26 (arguing that FTC jurisprudence has constituted a body of privacy law).

108. RING, <https://ring.com/> (last visited Aug. 4, 2022).

partnership with over 500 police departments,¹⁰⁹ Amazon allows police to contact Ring users and ask for surveillance footage. In exchange, Amazon provides departments with free or low-cost Ring devices to distribute to residents. Essentially, Amazon supports a government surveillance effort through consumer-grade products that are regulated primarily under consumer protection privacy laws. Previously, Amazon also sold to law enforcement agencies its “Rekognition network,” a facial recognition system that aimed to identify faces, objects, and scenes from images. Amazon announced a moratorium on police use of its facial recognition programs in 2020 and renewed that moratorium in 2021.¹¹⁰ Private-public partnerships like the Amazon-police department partnerships can create disparate harms. Surveillance programs may increase privacy violations for marginalized communities, and consumer protection law may not be enough to ward off those harms.

Privacy violations cause disparate harms to marginalized populations, particularly in an age of surveillance capitalism. This happens for a variety of reasons. Some groups may be less able to access or use new technologies. For example, poor people have less access than wealthy people to new consumer technologies or even internet access, despite the fact that they endure more surveillance and have less ability to fight back.¹¹¹ While it can be expensive to buy a new computer or afford stable housing with internet, it is free to be surveilled by the state. Due to disproportionate access to technology, any privacy protective benefits from new technologies could accrue less for people from marginalized populations.

Additionally, companies may have less of a market incentive to design products and services that cater to the needs of underprivileged consumers or minority consumers. This may occur simply because there are fewer people in minority groups and thus less of a market for products and services. Additionally, underprivileged consumers may be less able to spend on products and services, which decreases the market incentive for companies to build in privacy protections for those consumers. Thus, it is possible that fewer protections will be built into products to specifically address the privacy needs of those populations. Take stalkerware, for example. Stalkerware is the

109. Rani Molla, *Activists are pressuring lawmakers to stop Amazon Ring's police surveillance partnerships*, VOX (Oct 8, 2019, 7:00 AM), <https://www.vox.com/recode/2019/10/8/20903536/amazon-ring-doorbell-civil-rights-police-partnerships>.

110. Karen Weise, *Amazon indefinitely extends a moratorium on the police use of its facial recognition software*, N.Y. TIMES (May 18, 2021), <https://www.nytimes.com/2021/05/18/business/amazon-police-facial-recognition.html>.

111. See generally EUBANKS, *supra* note 24 (thoughtfully exploring multiple ways in which algorithmic systems discriminate against the poor).

colloquial term for digital applications that can help someone track and surveil another person, usually an intimate partner, disproportionately a woman or a girl.¹¹² However, technological design choices can stop developments like stalkerware,¹¹³ but without regulatory pressure, companies may have little incentive to solve these problems of privacy and equality.

Conversely, consumers with less means may be more attractive targets of exploitative data collecting practices. For example, Worldcoin, a cryptocurrency project, gained its first half a million users by targeting people in developing countries and promising cash and other financial benefits in exchange for users giving up their biometric data (body, face, and eye/iris scans).¹¹⁴ So far, Worldcoin has not produced any tangible products or services, and no cryptocurrency has emerged.

Consumer technologies today often rely on vast quantities of personal and other information that are then fed into a dangerous cyclical data ecosystem. This data ecosystem transfers individual consumer data to private and public parties that can then weaponize that data and use it for discriminatory purposes. For example, Clearview AI, a data aggregator, scraps personal images from social media websites and uses those images to power facial recognition algorithms that Clearview AI then sells to unknown entities, potentially helping empower authoritarian governments seeking to discriminate against political, religious, or ethnic minorities.¹¹⁵

In the absence of a strong federal privacy law, sectoral privacy laws leave many gaps where privacy violations can occur. This problem is not solely related to privacy but also to equality and civil rights. The sectoral privacy regime does not provide the regulatory push necessary to get companies to protect privacy equally for all individuals (or all consumers). A civil rights-

112. Christopher Parsons, Adam Molnar, Jakub Dalek, Jeffrey Knockel, Miles Kenyon, Bennett Haselton, Cynthia Khoo & Ron Deibert, *The Predator in Your Pocket: A Multidisciplinary Assessment of the Stalkerware Application Industry*, CITIZEN LAB (June 12, 2019), <https://citizenlab.ca/2019/06/the-predator-in-your-pocket-a-multidisciplinary-assessment-of-the-stalkerware-application-industry/>.

113. See generally Thomas E. Kadri, *Networks of Empathy*, 2020 UTAH L. REV. 1075 (2020) (arguing that technological and norms-based solutions can help fill in the gaps where regulation of digital abuse fails).

114. Eileen Guo and Adi Renaldi, “Deception, exploited workers, and cash handouts: How Worldcoin recruited its first half a million test users,” MIT Tech Review, <https://www.technologyreview.com/2022/04/06/1048981/worldcoin-cryptocurrency-biometrics-web3/> (April 6, 2022).

115. Caroline Haskins, Ryan Mac & Logan McDonald, *Clearview AI Wants to Sell its Facial Recognition Software to Authoritarian Regimes Around the World*, BUZZFEED NEWS, (Feb. 5, 2020, 8:15 PM), <https://www.buzzfeednews.com/article/carolinehaskins1/clearview-ai-facial-recognition-authoritarian-regimes-22>.

based understanding of privacy requires us to attempt to mitigate and resolve the distributed discrimination harms of the interconnected data ecosystem, including the downstream harms of data aggregation and use of data in artificial intelligence and machine learning systems.

C. THE PRIVACY TORTS

Privacy is an important value to individuals as citizens of a government, as consumers in an economy and as people in a society. Privacy protections in tort law can help shape the way our society recognizes privacy, including the way individuals interact with each other.¹¹⁶ Tort law has its limitations, and some of these limitations are particularly apparent when considering the cause of civil rights and the quest for equality.

In their landmark article conceptualizing the right of privacy in the United States, Samuel Warren and Louis Brandeis wrote that privacy was “the right to be let alone.”¹¹⁷ Critical scholars have suggested that this right to be let alone was, in particular, framed as such to protect the right for wealthy elites like Warren and Brandeis to conceal the private facts of their and their families’ lives from the public eye.¹¹⁸

Scholars like Amy Gajda, Anita L. Allen, and Erin Mack have written about how the right to privacy developed as a right for a privileged few (those who may have found themselves the subject of press gossip),¹¹⁹ perhaps with a

116. See generally Robert C. Post, *The Social Foundations of Privacy: Community and Self in the Common Law Tort*, 77 CALIF. L. REV. 957 (1989) (arguing that the common law tort of invasion of privacy safeguards social and community norms).

117. Warren & Brandeis, *supra* note 19, at 193.

118. See Anita L. Allen & Erin Mack, *How Privacy Got Its Gender*, 10 N. ILL. U. L. REV. 441, 457 (1991) (“Personal experiences with unwanted publicity concerning his Boston Brahmin family’s social life may have prompted Warren to coauthor the famous article.”); Samantha Barbas, *Saving Privacy from History*, 61 DEPAUL L. REV. 973, 983 (2012) (“Warren was incensed at finding details of his family’s home life and social affairs spread on the society pages of several newspapers. More broadly, the authors were outraged by the new trend of invasive news reporting and what they considered to be the unwarranted and tasteless depiction of private life in the press.”); cf. Dorothy J. Glancy, *The Invention of the Right to Privacy*, 21 ARIZ. L. REV. 1, 9 (1979) (arguing that Warren and Brandeis did not bow to elitism in their framing of the right to privacy, particularly in comparison to E. L. Godkin, who specifically denoted particular rights to privacy for different categories of people in his contemporaneous article, *The Rights of the Citizen: To His Own Reputation*, 8 SCRIBNER’S MAG. 58, 65 (1890)).

119. See generally Amy Gajda, *What if Samuel D. Warren Hadn’t Married a Senator’s Daughter?: Uncovering the Press Coverage that Led to “The Right to Privacy,”* 2008 MICH. STATE L. REV. 35, (2007) (examining contemporary news coverage of the Warren family that may have inspired Warren to coauthor *The Right to Privacy*, *supra* note 19); Allen & Mack, *supra* note 118, at 456 (“Overwrought by today’s standards, the Warren and Brandeis article was a lofty defense of values of affluence and gentility.”).

special eye toward enshrining in it certain values including “traditional norms of female modesty.”¹²⁰ Charles E. Colman has suggested that Warren may have been motivated to frame privacy in the way that he did out of concern for the privacy of his gay brother, Ned, and out of fear that Ned’s secrets would be exposed in a dangerous manner.¹²¹

This is all to say that even the origins of privacy are not devoid of considerations of equality and what many now call civil rights. Even the absence of such considerations in the original framing can be useful in understanding the limitations of privacy torts today and why reform is necessary to protect both privacy and civil rights in social interactions between individuals.

In Warren and Brandeis’s time, gossip society pages in print newspapers may have been among the greatest threats to a person’s reputation. However, today, with the explosion of the internet and online communications, there are many more threats to reputation and privacy. For example, Warren and Brandeis likely did not foresee deepfakes¹²² or nonconsensual sexual imagery. However, today, we live in a new world, where those harms exist, and those technologically driven privacy harms can impact a person’s access to equality and civil rights.

Today’s contemporary understanding of privacy in tort law undeniably also reflects William Prosser’s influential formalization of privacy torts in the *Second Restatement of Torts*¹²³ and elsewhere.¹²⁴ Prosser conceptualized privacy as four distinct torts: the right against reasonable intrusion, the right to publicity, the right against appropriation, and the right against publicity that places one in a false light.¹²⁵ Although the privacy torts can be useful in some cases, tort law has its limits in protecting privacy.¹²⁶

120. Allen & Mack, *supra* note 118, at 444.

121. Charles E. Colman, Comment, *About Ned*, 129 HARV. L. REV. F. 128, 151 (arguing that Warren may have been motivated to coauthor *The Right to Privacy* due to a desire to protect the privacy of his gay brother Ned).

122. See Danielle K. Citron & Robert Chesney, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753 (2019) (explaining privacy, democracy, and national security challenges related to deepfakes).

123. See generally Restatement (Second) of Torts § 652 (Am. L. Inst. 1975).

124. See generally e.g., Danielle Keats Citron, *Mainstreaming Privacy Torts*, 98 CALIF. L. REV. 1805, (2010) (discussing in part the lasting influence of Prosser’s conceptualization of privacy torts).

125. See Restatement (Second) of Torts.

126. See generally, e.g., Citron, *supra* note 124, at 1805; Neil M. Richards, *The Limits of Tort Privacy*, 9 J. TELECOMM. & HIGH TECH. L. 357 (2011); Patricia Sánchez Abril, *A (My)space of One’s Own: On Privacy and Online Social Networks*, 6 NW. J. TECH. & INTELL. PROP. 73 (2007); Barbas, *supra* note 118, at 973.

Privacy torts provide some legal basis for individuals who wish to use litigation to enforce their privacy rights. However, privacy torts do not account for inequality in privacy violations and protections. There will always be some who have more access to protection than others and more ability to litigate to enforce their rights. This power discrepancy is fundamentally antithetical to privacy as a civil right.

For example, in arguing for the conceptualization of the cyber civil rights discussed earlier in Section III.C, Danielle Keats Citron has noted the failings of tort law in protecting victims of cyber harassment, cyber stalking, and nonconsensual sharing of intimate imagery (“revenge porn”).¹²⁷ According to Citron, “Law’s shortcomings have made combating cyber harassment difficult. Tort remedies for defamation, intentional infliction of emotional distress, and privacy invasions exist only in theory for some victims, due to the high cost of litigation and the absence of specific privacy protections.”¹²⁸ Not only is the high cost of litigation a factor—the financial cost as well as the emotional and psychological cost—but the distributed and often anonymous nature of the internet may make it difficult to find and stop the perpetrators.

In any system of law, it is an unfortunate reality that those with greater economic means are often better able to pursue their legal rights in court. For privacy and tort law, this inequality exists here as well. As Scott Skinner-Thompson argues when discussing the inequality in applications of privacy torts, wealthy or famous people have greater ability to protect their tort interests, both due to having more means to sue and also due to a greater leeway afforded to them by unequal application of privacy tort law to public figures.¹²⁹ Thus, a tort law understanding of privacy that does not include civil rights can lead to further inequities for people from already marginalized populations.

V. RECOMMENDATIONS

To protect privacy and civil rights, and privacy as a civil right, it is important to consider equality and privacy as values that go hand in hand. Across different sectors of privacy regulation, the law can and should incorporate an understanding of the civil rights, antidiscrimination, and anti-subordination impacts of privacy laws.

We should understand privacy both as a civil right and as a core component to creating the conditions that allow civil rights to flourish. Not

127. See generally Citron, *supra* note 71.

128. CITRON, *supra* note 72, at 24.

129. SKINNER-THOMPSON, *supra* note 23, at 195.

only do individuals deserve equal privacy protections, but they also deserve protections for the privacy necessary to advocate for equality. Traditional civil rights laws protect equal access to spaces and opportunities. Cyber civil rights protect these rights in the online world. Now it is time to understand how to protect both privacy and civil rights in a world where technology blurs the line between online and offline and where technologically mediated civil rights violations in the offline world can harm individuals. Where access to privacy protection is unequally distributed in a discriminatory fashion or where certain privacy practices (invasions of privacy or protections of privacy) discriminate against individuals from protected classes, the law ought to view this as a civil rights violation.

Constitutionally, courts should begin to understand privacy as integral to both due process and equal protection. That is, there may be policies and practices that violate privacy in a way that violates an individual's due process and equal protection rights as well. The category of claims that should give rise to a combined privacy and civil rights interest, constitutionally, should not be limited to only laws against gay marriage and abortion. Laws and practices that promote surveillance, mandate the use of biased algorithmic assessments, and allow for gendered harms related to cyber stalking, should also be considered unconstitutional based on due process and equal protection. Individuals should be able to claim a constitutional right to privacy under the Equal Protection Clause, recognizing that privacy has never been awarded equally to all people across society.

Tort laws for privacy merit reform as well. There are clear limitations on the ability for privacy torts to truly protect the privacy rights of all individuals, particularly given the disparity in ability to bring or prevail on tort privacy claims and the difficulty of traditional privacy tort doctrine in applying to modern technologies like social media platforms. We should recognize the equal access interests for plaintiffs in privacy tort actions and understand the civil rights implications of privacy torts. Where possible, civil procedure can also be reformed to balance the disparity of power between plaintiffs who belong to protected classes (and other marginalized groups, e.g., the poor) and those who do not. In short, equal protection principles should be applied to the protection of privacy in tort law to protect both privacy and civil rights.

State and sectoral privacy laws, particularly those that perceive of privacy as a consumer protection right, must take into account civil rights as well. For example, Ifeoma Ajunwa has called for Congress to add a disparate impact clause into the Genetic Information Nondiscrimination Act to fully address

the potential for genetic discrimination, particularly the outsized impact on certain protected classes.¹³⁰ Congress should heed this call.

The United States should pass a federal privacy law to fill in the gaps of the sectoral privacy regime, and this federal privacy law must be written with civil rights in mind. A federal privacy law must take into account the civil rights implications of privacy. The law must recognize that privacy risks and harms are unequally distributed. This must be part of the framing as a signaling move to enshrine privacy and civil rights into our legal system.

U.S. jurisprudence tends to shy away from stipulating positive rights for individuals. However, critically, one space where this may be less of a factor is concerning civil rights. Antidiscrimination laws often amount to laws that impose positive obligations on state actors to create conditions that make equality possible.¹³¹ For example, in mandating compliance with the Americans with Disabilities Act (ADA) standards, the government arguably creates positive obligations for entities that offer public accommodations to create accessible spaces for accommodation, even in online spaces.

If we are to truly believe in the importance of privacy and the rights of all individuals to have equal access to privacy and equal protection against privacy violations, then we ought to protect privacy as a civil right. Individuals should be able to raise civil rights claims when they believe they are the subject of a privacy violation that amounts to discrimination on the basis of a protected characteristic, even when that discrimination occurs online or in a hybrid context. Individuals should also have civil rights claims against unfair and unequal privacy violations, like the disproportionate use of surveillance technology against marginalized populations. The civil right of privacy should be codified into a federal privacy law or an extensions of current privacy laws or civil rights laws.

A civil rights-oriented right to privacy should conceptualize privacy as a positive right. To do so, legislators can borrow from international conceptions of privacy rights, including conceptions of privacy as a primarily dignitary interest (as in Article 7 of the Charter of Fundamental Rights of the European Union¹³²) and privacy as a fundamental human right (as in Article 12 of the

130. Ifoema Ajunwa, *Genetic Data and Civil Rights*, 51 HARV. C.R.-C.L. L. REV. 75, 100 (2016).

131. Of course, one could say that a positive obligation to create conditions that make equality possible is functionally equivalent to a negative obligation against creating conditions that make equality impossible. To which I say, oh well.

132. Charter of Fundamental Rights of the European Union, art. 7, 2000 O.J. (C 364) 1.

Universal Declaration of Human Rights¹³³ and Article 17 of the International Covenant on Civil and Political Rights).¹³⁴ Viewed in this manner, privacy becomes not a right to be let alone or a right against government intrusion, but a positive right for individuals that obligates the government to provide the conditions that make privacy possible. Centering equality and civil rights into our privacy laws can also create a strong signaling factor to other nations, shifting the global norms for privacy in a way that moves us toward a more open, democratic, and equal future.

VI. CONCLUSION

Before final publication of this article, the Supreme Court released its decision in the controversial case, *Dobbs v. Jackson Women's Health Organization*.¹³⁵ This decision effectively overturned *Roe v. Wade*, the monumental case that legalized the right to abortion nationwide.¹³⁶ The *Dobbs* decision has upset decades of national legal protection for the right to abortion and, potentially, decades of precedent on the scope of a person's right to privacy as well as the limits of equal protection as related to substantive due process.

Not only has the *Dobbs* decision reinforced the need for greater privacy protections, generally, but this recent ruling has also cast more light on the impact that technology can have on privacy rights.¹³⁷ In states where abortion is or may soon be criminalized, a person's digital data may potentially be used against them in court. For example, location data from cell phones could be used as evidence to show that a person visited an abortion clinic. Emails and messaging data could include conversations about seeking or providing abortions, which could implicate a person in illegal actions, if in a state where abortions are illegal.

In post-*Roe* America, digital privacy, health privacy, bodily autonomy, and equal protection are more intertwined than ever. Losing the nationwide right to abortion creates harm that disproportionately affects already marginalized,

133. Universal Declaration of Human Rights, art. 12, U.N. GAOR, Supp. No. 16, at 52, U.N. Doc. A/6316 (1948) ("No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks.")

134. See International Covenant on Civil and Political Rights art. 17, Dec. 16, 1966, 999 U.N.T.S. 171 S. Treaty Doc. 95-20.

135. *Dobbs v. Jackson Women's Health Org.*, 597 U. S. ____ (2022).

136. *Roe v. Wade*, 410 U.S. 113 (1973).

137. Tiffany Li, *In post-Roe America, your cell phone is now a reproductive privacy risk*, MSNBC DAILY, <https://www.msnbc.com/opinion/msnbc-opinion/states-abortion-bans-can-weaponize-your-own-data-against-you-n1296591> (June 25, 2022).

vulnerable groups, including Black women, who are more likely to live in states that already ban or may soon ban abortion, more likely to seek abortion care, and more likely to die in childbirth.¹³⁸ The right to abortion is a privacy right, and it should also be considered a civil right. As with many other forms of privacy violations, restrictions on the right to abortion amount to restrictions on the right to privacy that are unequally applied and disparate in impact – violations of civil rights.

While civil rights law in the United States has its limitations, conceptually situating privacy rights in a civil rights context can help us remedy, or at least expose, some of the inequities in access to privacy protections. Protecting privacy as a civil right serves the dual function of protecting the civil right of privacy and protecting the ability for privacy to support and enhance other civil rights. Furthermore, it is critical to consider privacy as a fundamental tool for protecting civil rights because our new technologically motivated world demands it.

As citizens of the future, we must recognize that the ever-changing nature of new technologies will continue to shape our conceptions of privacy and civil rights. Privacy should be considered a civil right, particularly because our new, connected world has changed the way we understand our data and ourselves. No longer is it enough to protect equal access to physical spaces and equal rights in the offline world. Now, we must protect our privacy online, offline, and in situations where the line between cyber and physical space is blurred.

The right to privacy is core to civil rights and the fight for freedom and equality worldwide. There cannot be a just and equal society when the rights to privacy are not protected equally—when some are unable to access privacy protections afforded to them by law, while others are the subject of disproportionate privacy violations. Privacy and civil rights are inextricably intertwined, and the law and legal scholars must understand this complexity in order to better protect both core values for the future.

138. Nandita Bose, *Roe v Wade ruling disproportionately hurts Black women, experts say*, REUTERS, <https://www.reuters.com/world/us/roe-v-wade-ruling-disproportionately-hurts-black-women-experts-say-2022-06-27/> (June 27, 2022).

CONTENT MODERATION AS SURVEILLANCE

Hannah Bloch-Wehba[†]

ABSTRACT

Technology platforms are the new governments, and content moderation is the new law, or so goes a common refrain. As platforms increasingly turn toward new, automated mechanisms of enforcing their rules, the apparent power of the private sector seems only to grow. Yet beneath the surface lies a web of complex relationships between public and private authorities that call into question whether platforms truly possess such unilateral power. Law enforcement and police are exerting influence over platform content rules, giving governments a louder voice in supposedly “private” decisions. At the same time, law enforcement avails itself of the affordances of social media in detecting, investigating, and preventing crime.

This Article, prepared for a symposium dedicated to Joel Reidenberg’s germinal article *Lex Informatica*, untangles the relationship between content moderation and surveillance. Building on Reidenberg’s fundamental insights regarding the relationships between rules imposed by legal regimes and those imposed by technological design, the Article first traces how content moderation rules intersect with law enforcement, including through formal demands for information, informal relationships between platforms and law enforcement agencies, and the impact of end-to-end encryption. Second, it critically assesses the degree to which government involvement in content moderation actually tempers platform power. Rather than effective oversight and checking of private power, it contends, the emergent arrangements between platforms and law enforcement institutions foster mutual embeddedness and the entrenchment of private authority within public governance.

DOI: <https://doi.org/10.15779/Z389C6S202>

© 2021 Hannah Bloch-Wehba.

[†] Associate Professor, Texas A&M School of Law. My thanks to Kendra Albert, Julie Cohen, Ignacio Cofone, Caroline Mala Corbin, Rebecca Crootof, Angel Diaz, evelyn douek, Miriam Estrin, Nik Guggenberger, Thomas Kadri, Christina Koningisor, Rachel Levinson-Waldman, Przemek Palka, Christopher Reed, Alicia Solow-Niederman, Jennifer Urban, and participants at the Freedom of Expression Scholars Conference, WIPIP and Platgov for helpful comments on this project. I am very grateful to Joshua Frechette for excellent research assistance. Finally, the terrific editors at the *Berkeley Technology Law Journal*, including Loc Ho, Justine McCarthy Potter, Barbara Rówińska, and Dakota Sneed, who helped get this Article across the finish line. All mistakes are, of course, my own.

TABLE OF CONTENTS

I.	INTRODUCTION	1298
II.	POLICING’S INFLUENCE ON PLATFORMS	1303
	A. INTERMEDIARY PROTECTION AND PRIVATE GOVERNANCE	1304
	B. FORMAL INDEPENDENCE, INFORMAL ENTANGLEMENT	1307
	1. Terrorist Content.....	1307
	2. Sex Work.....	1310
	C. NEW INCENTIVES FOR PLATFORMS?	1313
III.	PLATFORMS’ INFLUENCE ON POLICING	1314
	A. SHAPING LAW ENFORCEMENT THROUGH TECHNOLOGY	1315
	1. Compelled Disclosure.....	1315
	2. “Open Source” Investigations.....	1318
	3. Deputizing Users	1320
	4. Resistance Through Design	1323
	B. SHAPING LAW ENFORCEMENT THROUGH PLATFORM POLICY	1326
	C. VOLUNTARY PRIVATE-PUBLIC SURVEILLANCE ARRANGEMENTS	1328
IV.	IMPLICATIONS FOR CRIMINAL PROCEDURE	1331
	A. THE EMERGENCE OF NEW FORMS OF DISCLOSURE	1331
	B. NEW INVESTIGATIVE METHODS	1333
	C. DESIGN AND LEGAL IMMUNITY	1337
V.	CONCLUSION	1339

I. INTRODUCTION

In September 2020, after a summer of uprisings against police violence, a series of wildfires broke out in the Northwest. It didn’t take long for rumors that the fires had been started by antifa activists, or by the Proud Boys, to start spreading on social media. Soon, vigilantes set up roadblocks, searching for the responsible parties and, in the process, obstructing traffic and heightening tensions. Law enforcement agencies, tasked with enforcing evacuation orders, grew increasingly concerned about viral misinformation making their jobs even harder.¹

1. Dennis Romero, *Facebook to Take Down False Reports of Antifa Arson in Oregon*, NBC NEWS (Sept. 13, 2020, 2:58 AM), <https://www.nbcnews.com/tech/tech-news/facebook-take-down-false-reports-antifa-arson-oregon-n1239966>.

After first working to attach misinformation “warning labels” to the posts, Facebook ultimately announced that it would delete the posts altogether.² Facebook’s action was welcome, but puzzling to some. After a year of epic failures in addressing misinformation about public health, elections, and social movements, why did Facebook act so quickly—and so aggressively—in shutting down misinformation about the Oregon wildfires? This Article proposes a potential answer: law enforcement’s assertion of its own demands and needs shaped Facebook’s content moderation rules and affected Facebook’s response to crisis.

This Article suggests that law enforcement’s impact on content governance is not sporadic or fleeting. Policing is, instead, a durable influence on the rules, standards, and technical processes by which platforms govern their communities. Nor is this influence limited to high profile examples of unlawful speech, such as terrorism, incitement of violence, or sex trafficking. In more quotidian contexts, platforms also play a crucial role as intermediaries in evidence-gathering processes.³ As police increasingly depend upon digital evidence in investigating and prosecuting crime, content governance strategies also shape the kinds of data that are germane to investigations and affect how law enforcement does its job.⁴

This commingling of public and private authority raises conceptual questions about the nature of content and data governance. While a robust literature considers how and why platforms have developed “community standards” by which they govern user behavior in online spaces, the “private” character of these standards and rules is often taken for granted.⁵ Indeed, platforms are often described as governments in their own right, equally powerful and sovereign as the states in which they are headquartered.⁶ The structure of intermediary liability law reaffirms this conception of platform

2. Reuters Staff, *Facebook Removes Posts Linking Oregon Wildfires to Activist Groups*, REUTERS (Sept. 13, 2020, 3:19 AM), <https://www.reuters.com/article/us-usa-wildfires-facebook-idUSKBN264013>.

3. *See infra* Part III.A.

4. *See infra* Part III.B.

5. *See, e.g.*, Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598 (2018); Kyle Langvardt, *Regulating Online Content Moderation*, 106 GEO. L.J. 1353 (2017); NICOLAS P. SUZOR, *LAWLESS: THE SECRET RULES THAT GOVERN OUR DIGITAL LIVES* 11 (2019) (“The legal reality is that social media platforms belong to the companies that create them, and they have almost absolute power over how they are run.”).

6. Kristen E. Eichensehr, *Digital Switzerlands*, 167 U. PA. L. REV. 665, 672 (2019); *see also* JULIE E. COHEN, *BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM* 122 (2019) (noting that platforms “have worked to position themselves as both essential partners and competing sovereigns in the quest to instantiate states of exception algorithmically”).

governance as “private.” At least under U.S. law (for now), platforms are largely immune from liability for hosting even unlawful user-generated content, leading scholars to describe the voluntary mechanisms they enforce as a category of private regulation adopted without legal obligation.

In fact, however, the purportedly private rules of content moderation emerge and operate within a political context in which law enforcement acts as a particularly powerful stakeholder. For example, law enforcement has encouraged platforms to adopt more stringent rules on certain categories of harmful content, such as child sexual abuse imagery (CSAM) or violent rap music in the UK.⁷ In Europe, police agencies have formed special “internet referral units” to report and flag violations of platforms’ content rules for takedown.⁸ Platforms’ private decision-making thus provides a new avenue for law enforcement to regulate the public sphere.⁹ As platform firms turn to automation and artificial intelligence to scale up their efforts to address harmful online content, the technical infrastructures of content moderation increasingly reflect government influence.¹⁰

Just as law enforcement seeks expanded influence over platforms’ private decision-making, the processes and technical affordances of content governance also affect and shape law enforcement investigations in more mundane contexts. Police rely on social media to identify purported gang members, generate investigative leads, map networks and associations, and monitor activity by the public.¹¹ The prevalence of social media as an

7. See *infra* Part II.A.

8. Brian Chang, *From Internet Referral Units to International Agreements: Censorship of the Internet by the UK and EU*, 49 COLUM. HUM. RTS. L. REV. 114, 120–22 (2017); Hannah Bloch-Wehba, *Global Platform Governance: Private Power in the Shadow of the State*, 72 SMU L. REV. 27, 45–46 (2019).

9. Susan Benesch, *But Facebook’s Not a Country: How to Interpret Human Rights Law for Social Media Companies*, 38 YALE J. ON REGUL. BULL. 86, 99 (2020) (“Company content moderation is also used as a means for states to carry out silent and invisible censorship.”); see also Bloch-Wehba, *supra* note 8, at 45–46 (distinguishing between legal takedown orders and the expanded global sweep of takedowns under platforms’ internal terms of service).

10. Hannah Bloch-Wehba, *Automation in Moderation*, 53 CORNELL INT’L L.J. 41, 69–70 (2020).

11. See Katherine J. Strandburg, *Freedom of Association in a Networked World: First Amendment Regulation of Relational Surveillance*, 49 B.C. L. REV. 741, 743–44 (2008) (describing how law enforcement began to use communications traffic data to map social relationships and group memberships, and naming these strategies “relational surveillance”); Desmond Upton Patton, Douglas-Wade Brunton, Andrea Dixon, Reuben Jonathan Miller, Patrick Leonard & Rose Hackman, *Stop and Frisk Online: Theorizing Everyday Racism in Digital Policing in the Use of Social Media for Identification of Criminal Conduct and Associations*, 3 SOCIAL MEDIA + SOCIETY 1 (2017) (describing the use of social media information in gang databases); Megan Behrman, *When Gangs Go Viral: Using Social Media and Surveillance Cameras to Enhance Gang Databases*, 29 HARV. J. L. & TECH. 315 (2015); Keegan Stephan, *Conspiracy: Contemporary Gang*

investigative tool also makes investigations, to some degree, reliant on platforms' own decisions about what content-related behaviors to permit or forbid. For instance, users' ability to delete posts, photos, videos, emails, and messages has prompted law enforcement agencies to procure new tools to scrape and retain user data.¹² Ironically, as platforms have cracked down on certain types of unlawful content, they have arguably made law enforcement's jobs in ferreting out unlawful activity that much more difficult.¹³

The chief goal of this Article is to illuminate the close relationship between platforms and police by examining how content-related decision-making within private platforms can advance or inhibit law enforcement surveillance practices. In so doing, I bring together two distinct bodies of scholarship. The first emphasizes platforms' roles as private guarantors of free expression and views government pressures on content-related rules as a toxic form of "jawboning" or collateral censorship through which the government seeks to regulate the public sphere indirectly when it could not do so directly.¹⁴ The second examines how government can compel disclosure or otherwise extract information about users from social media and the role of internet platforms in accommodating, facilitating, and resisting those demands.¹⁵ As Joel

Policing and Prosecutions, 40 CARDOZO L. REV. 991, 1021 (2018) ("The NYPD has admitted that communicating with the wrong person on social media is enough to get someone placed on a gang database . . ."); Sam Biddle, *Police Surveilled George Floyd Protests with Help from Twitter-Affiliated Startup Dataminr*, THE INTERCEPT (July 9, 2020, 8:00 PM), <https://theintercept.com/2020/07/09/twitter-dataminr-police-spy-surveillance-black-lives-matter-protests/>; Bill Dries, *Police Documents Show Protest Spreadsheet and Fear of 'Radical'*, MEMPHIS DAILY NEWS (July 31, 2018), <https://www.memphisdailynews.com/news/2018/jul/31/police-documents-show-protest-spreadsheet-and-fear-of-radicals//print>.

12. Kate Knibbs, *The Race to Preserve the DC Mob's Digital Traces*, WIRED (Jan. 7, 2021, 5:40 PM), <https://www.wired.com/story/archive-social-media-footage-pro-trump/>; see *infra* text accompanying notes 122–129.

13. See, e.g., Mike Masnick, *More Police Admitting That FOSTA/SESTA Has Made It Much More Difficult to Catch Pimps and Traffickers*, TECHDIRT, <https://www.techdirt.com/articles/20180705/01033440176/more-police-admitting-that-fosta-sesta-has-made-it-much-more-difficult-to-catch-pimps-traffickers.shtml> (last visited Mar. 25, 2021).

14. See, e.g., Seth F. Kreimer, *Censorship by Proxy: The First Amendment, Internet Intermediaries, and the Problem of the Weakest Link*, 155 U. PA. L. REV. 11 (2006); Jack M. Balkin, *Old-School/New-School Speech Regulation*, 127 HARV. L. REV. 2296 (2013); Derek E. Bambauer, *Against Jawboning*, 100 MINN. L. REV. 51 (2015); Danielle Keats Citron, *Extremist Speech, Compelled Conformity, and Censorship Creep*, 93 NOTRE DAME L. REV. 1035 (2017).

15. Daniel J. Solove, *The First Amendment as Criminal Procedure*, 82 N.Y.U. L. REV. 112 (2007); Katherine J. Strandburg, *Home, Home on the Web and Other Fourth Amendment Implications of Technosocial Change Special Feature: Cyberlaw*, 70 MD. L. REV. 614 (2010–11); Jonathan Manes, *Online Service Providers and Surveillance Law Transparency*, 125 YALE L.J. F. 343 (2015–16); Alan Z. Rozenshtein, *Surveillance Intermediaries*, 70 STAN. L. REV. 99 (2018); Rebecca Wexler, *Privacy as Privilege: The Stored Communications Act and Internet Evidence*, 134 HARV. L. REV. 2721 (2021).

Reidenberg suggested in *Lex Informatica*, these two conceptions address two siloed visions of the role of platforms in constituting and governing the public sphere.¹⁶

This Article makes three contributions. First, it complicates existing narratives about the respective roles of social media platforms and law enforcement agencies regarding effective policing. Second, the Article maps how law enforcement both influences and relies upon platform content governance. Although law enforcement seeks to influence *lex informatica*, the substantive, procedural, and technical rules of platforms also shape law enforcement itself. Finally, the Article examines the implications of this public-private cooperation for the law of criminal procedure. Understanding how (public) law enforcement and (private) platform rules mutually inform and co-constitute each other complicates the existing division in U.S. law between state and private action.¹⁷

The rest of the Article proceeds in three parts. Part II reviews how protections from intermediary liability encouraged the development of private platform governance through technology, even as law enforcement needs remained a powerful influence on firms. Today, contemporary debates over changes to intermediary liability rules highlight the risk that new regulations might promote state censorship laundered through private actors.¹⁸ Even without legal change, however, recent decisions by payment processors and social media platforms reflect the continuing influence of law enforcement even in “private” domains, as Part II.B recounts.

Part III develops the idea that the emergence of online commerce and communication has fundamentally reshaped law enforcement investigative practices. Part III further illustrates that the technological affordances of platforms drive policing’s appetite for more data.¹⁹ And while platforms sometimes constrain policing through privately developed policy, the mechanisms of private governance also advance law enforcement strategies.²⁰ The result is that the technological modalities of governing online content also

16. Joel R. Reidenberg, *Lex Informatica: The Formulation of Information Policy Rules Through Technology*, 76 TEX. L. REV. 553, 554 (1997) (distinguishing between “[t]he treatment of content” and “the treatment of personal information”).

17. Cf. Kiel Brennan-Marquez, *The Constitutional Limits of Private Surveillance*, 66 U. KAN. L. REV. 485, 490 (2018) (arguing that whether surveillance is conducted by state or private actors may not matter if it “threaten[s] the integrity of social life”).

18. See *infra* Part II.C.

19. See *infra* Part III.A.

20. See *infra* Part III.B.

have increasing resonance for law enforcement investigations despite their “private” character.²¹

Part IV considers the implications of the increasing enmeshment of private platforms and law enforcement for the law of criminal procedure. The turn toward automated modalities of content governance will create new types and sources of information relevant to new kinds of investigations.²² Yet more extensive collaboration between law enforcement and platforms will raise difficult questions about how best to vindicate important accountability and transparency values when private firms play an increasingly significant role in facilitating public functions.

II. POLICING’S INFLUENCE ON PLATFORMS

People use social media to keep up with their friends and family, watch music and cooking videos, and consume news and political commentary. But social media is also home to a slew of unlawful content. For example, YouTube hosts videos that infringe copyright,²³ Facebook Marketplace features posts advertising drugs, sex, and guns,²⁴ and Twitter is home to coded posts advertising child sexual abuse imagery.²⁵ Yet under Section 230 of the Communications Decency Act of 1996, none of these sites can be held liable for hosting content that violates the law, with only a few exceptions.²⁶ This Part explores how, in spite of existing protections insulating them from liability, platforms have developed many formal and informal mechanisms for advancing law enforcement interests. Although not uniform, these mechanisms illustrate that platforms frequently accommodate law

21. *Id.*

22. *See infra* Part III.

23. *See* Kristelia García, *Monetizing Infringement*, 54 U.C. DAVIS L. REV. 265, 286 (2020) (describing how the scale of copyright infringement on YouTube led the platform to develop Content ID, an automated content screening tool).

24. Parmy Olson & Zusha Elinson, *Gun Sellers are Sneaking Onto Facebook’s Booming Secondhand Marketplace*, WALL ST. J. (Aug. 20, 2019, 5:57 PM), <https://www.wsj.com/articles/gun-sellers-are-sneaking-onto-facebooks-booming-secondhand-marketplace-11566315198>; Ananya Bhattacharya, *Facebook’s New Marketplace is Already Flooded with Illegal Guns, Drugs, Sex, and Wildlife*, QUARTZ, <https://qz.com/799943/facebook-fb-new-marketplace-is-already-flooded-with-illegal-guns-drugs-sex-and-wildlife/> (last visited July 15, 2021).

25. Olivia Solon, *Child Sexual Abuse Images and Online Exploitation Surge During Pandemic*, NBC NEWS (Apr. 23, 2020, 9:01 PM EST), <https://www.nbcnews.com/tech/tech-news/child-sexual-abuse-images-online-exploitation-surge-during-pandemic-n1190506>.

26. 47 U.S.C. § 230; *see also* Danielle Keats Citron & Benjamin Wittes, *The Internet will Not Break: Denying Bad Samaritans Sec. 230 Immunity*, 86 FORDHAM L. REV. 401, 403 (2017) (describing how Section 230 immunity has been extended to “immunize platforms dedicated to abuse and others that deliberately host users’ illegal activities”).

enforcement needs (although, at times, they also resist law enforcement demands).

A. INTERMEDIARY PROTECTION AND PRIVATE GOVERNANCE

While Section 230(c)(1) immunizes platforms from liability for most content posted by users,²⁷ Section 230(c)(2)'s "Good Samaritan" provision also protects providers that restrict "obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable" content.²⁸ The Good Samaritan provision explicitly immunizes online platforms that choose to edit or curate content in ways that would violate the First Amendment if done by the government itself.²⁹

The result is that Section 230 ranks "among the most important protections of free expression in the United States in the digital age."³⁰ It also set the stage for the emergence and growth of what Joel Reidenberg called "lex informatica."³¹ Section 230's Good Samaritan provision created breathing room within which self-regulation and private standard setting became the norm.³² Without the obligation to monitor, filter, or block content, intermediaries nonetheless began to do so, developing both new rules to shape their communities and new enforcement technologies.³³ The example of spam filtering is illustrative: facing a flood of unsolicited commercial advertising,

27. 47 U.S.C. § 230(c)(1) ("No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.").

28. 47 U.S.C. § 230(c).

29. *See, e.g.*, United States v. Stevens, 559 U.S. 460, 468 (2010) (concluding that depictions of animal cruelty do not fall into a category of speech that is unprotected by the First Amendment); FACEBOOK COMMUNITY STANDARDS, *Coordinating Harm and Publicizing Crime*, https://www.facebook.com/communitystandards/coordinating_harm_publicizing_crime/ (banning content "depicting, admitting to or promoting[,] [a]cts of physical harm against animals"); *see also* Domen v. Vimeo, Inc., 991 F.3d 66, 68 (2021) (reasoning that Section 230(c)(2) protects online video hosting service from liability when it deletes a user account that violates its policy against the promotion of conversion therapy).

30. Balkin, *supra* note 14, at 2313.

31. Reidenberg, *supra* note 16, at 555; *see also* Jack L. Goldsmith, *Against Cyberanarchy*, 65 U. CHI. L. REV. 1199, 1215–16 (1998) ("Private legal ordering thus has the potential to resolve many, but not all, of the challenges posed by multijurisdictional cyberspace activity.").

32. Reidenberg, *supra* note 16, at 583 ("Law may encourage the development of Lex Informatica by imposing liability on various network actors, and law may provide immunity or safe harbors for implementation of technical rules."); *see also* Klonick, *supra* note 5, at 1603–04 (linking private governance to legal immunity).

33. *See, e.g.*, Bloch-Wehba, *supra* note 10, at 52 (describing the development of spam filtering).

platforms developed anti-spam rules, protocols, and software to filter out unwanted ads.³⁴

But while protections for intermediaries allowed “private” regulation to flourish, formal immunity from liability does not equate to immunity from government pressure.³⁵ Even with Section 230’s liability shield intact, government agencies often engage in efforts to coerce, compel, or convince intermediaries to take down harmful content or provide information about the users who posted it.³⁶ These dynamics may transform online intermediaries into engines of unaccountable private censorship. Scholars of free speech worry that in controversial cases, the government might pressure online intermediaries to go along with the state’s own preferences for online speech, a form of “soft censorship” or “jawboning.”³⁷

Take the example of drill music, a genre of rap pioneered on Chicago’s South Side and popular in its own right in the United Kingdom.³⁸ To earn a living, drill artists rely on social media to distribute music videos that contain “morally charged caricatures of themselves,” replete with guns, violent lyrics, and drugs.³⁹ But drill music’s violent content and links to offline crime have also earned it the attention of law enforcement.⁴⁰ During a rise in violent crime

34. *Id.* at 55 (describing how platforms turned to automated technology to scale the fight against spam but adopted different definitions of prohibited spam activity).

35. Balkin, *supra* note 14, at 2314 (“What a system of intermediary immunities and safe harbors does not protect, however, constitutes a system of intermediary liability and, hence, of potential collateral censorship.”); see also Chris Montgomery, *Can Brandenburg v. Ohio Survive the Internet and the Age of Terrorism?: The Secret Weakening of a Venerable Doctrine*, 70 OHIO ST. L.J. 141, 168–78 (2009) (describing how law enforcement has encouraged voluntary action by ISPs and communications service providers).

36. Jonathan Zittrain, *Internet Points of Control*, 44 B.C. L. REV. 653, 674 (2003) (describing Pennsylvania law that required internet service providers to remove or block access to child pornography within five business days); Bambauer, *supra* note 14, at 67–68 (recounting how, under pressure from law enforcement institutions, states adopted laws meant to hold Backpage.com liable for posts submitted by users, knowing that those laws were likely unenforceable).

37. Derek E. Bambauer, *Orwell’s Armchair*, 79 U. CHI. L. REV. 863, 905 (2012) (describing process-oriented problems with “soft censorship”); Bambauer, *supra* note 14, at 61 (defining “jawboning” as “enforcement through informal channels, where the underlying authority is in doubt”).

38. Lambros Fatsis, *Policing the Beats: The Criminalisation of UK drill and Grime Music by the London Metropolitan Police*, 67 SOCIO. REV. 1300, 1302 (2019); Ben Beaumont-Thomas, *Is UK Drill Music Really Behind London’s Wave of Violent Crime?* (Apr. 9, 2018), <http://www.theguardian.com/music/2018/apr/09/uk-drill-music-london-wave-violent-crime>.

39. FORREST STUART, *BALLAD OF THE BULLET: GANGS, DRILL MUSIC, AND THE POWER OF ONLINE INFAMY* 6 (2020).

40. *YouTube Must Crack Down on Videos Pushing Violence & Knife Crime*, MAYOR OF LONDON (Aug. 7, 2017), <https://www.london.gov.uk/city-hall-blog/youtube-must-crack-down-videos-pushing-violence-knife-crime>.

in London, Cressida Dick, the Commissioner of the Metropolitan Police, began to pressure social media platforms to take down UK drill videos, citing their relationship to knife crime.⁴¹ In response, YouTube began aggressively taking down drill videos pursuant to police requests and developed special policies “specifically to help tackle videos related to knife crime in the UK.”⁴² YouTube also embraced close relationships with the police, publicizing its “dedicated process for the police to flag videos directly to [YouTube’s] teams.”⁴³ From the police perspective, stemming the dissemination of drill videos was only one part of a multiprong strategy. Police also obtained a “criminal behavior order” enjoining five people from “mentioning death or injury” in their online videos.⁴⁴ The Metropolitan Police also announced that it was indexing and tracking an extensive list of drill videos.⁴⁵

Sometimes, however, platforms push back against government demands. Consider, for example, the infamous “Innocence of Muslims” video, an Islamophobic “film” that sparked violent protests across the world and reportedly led to the attack on the U.S. consulate in Benghazi, Libya.⁴⁶ As violence spread, the White House reportedly called YouTube to ask the firm

41. *Met Police Chief Calls on YouTube to Take Down Drill Music to Curb Gang Crime*, LBC (May 18, 2018, 9:17 AM), <https://www.lbc.co.uk/radio/presenters/nick-ferrari/met-police-chief-calls-on-youtube-drill-music/>.

42. Lizzie Dearden, *Police Targeting Drill Music Videos in Controversial Crackdown on Social Media That ‘Incites Violence’*, THE INDEPENDENT (May 29, 2018, 12:04 AM), <https://www.independent.co.uk/news/uk/crime/drill-music-stabbings-london-youtube-violence-police-knife-crime-gangs-a8373241.html>; Jim Connolly, *Home Secretary: ‘Sweep the Net, Take Down Knife-crime Posts’*, BBC NEWS: NEWSBEAT (Feb. 13, 2019), <https://www.bbc.com/news/newsbeat-47211631>; Ed Clowes, *For British Drill Stars, the Police are Listening Closely*, N.Y. TIMES (Jan. 11, 2021), <https://www.nytimes.com/2021/01/11/arts/music/digga-d-drill-music.html> (charting rise in YouTube’s takedown numbers).

43. Dearden, *supra* note 42.

44. *Ladbroke Grove Banned From Making ‘Violent Drill Music’*, BBC NEWS (June 15, 2018), <https://www.bbc.com/news/uk-england-london-44498231>; Lanre Bakare, *‘New Stop and Search’: Rappers Condemn Police Over Drill Bans*, THE GUARDIAN (June 14, 2019), <http://www.theguardian.com/music/2019/jun/14/rappers-konan-krept-condemn-police-criminalisation-of-drill> (describing how two rappers were sentenced to prison for breaching a gang injunction prohibiting them from performing violent lyrics).

45. Jim Edwards, *YouTube Deleted 130 Rap Videos to Help Police Fight Street Gangs Responsible for Thousands of Stabbings*, BUS. INSIDER (June 29, 2019, 12:52 PM), <https://www.businessinsider.com/uk-drill-rap-videos-banned-by-police-2019-6> (quoting Met police as saying that their database contained over 2,000 music videos, while they had filed only 154 takedown requests with YouTube).

46. Michael Joseph Gross, *The Making of The Innocence of Muslims: Cast Members Discuss the Film That Set Fire to the Arab World*, VANITY FAIR (Dec. 27, 2012), <https://www.vanityfair.com/culture/2012/12/making-of-innocence-of-muslims>.

to review whether the video complied with its terms of service.⁴⁷ President Obama told 60 Minutes that while “we believe in the First Amendment,” the film “is not representative of who we are and our values.”⁴⁸ Civil liberties advocates chafed at the White House’s use of quasi-official channels to pressure YouTube to take down the offensive but lawful video, and YouTube ultimately resisted the calls to take the video down for a U.S. audience.⁴⁹

These two illustrations demonstrate YouTube’s power to either facilitate or obstruct law enforcement priorities. In the “Innocence of Muslims” case, YouTube’s own content-related policies led it to resist the White House’s encouragement to take down the video. Across the pond, however, YouTube created new content-related rules at law enforcement’s behest, offering itself as a vital partner to police. In both cases, YouTube’s decisions were formally voluntary, free of government coercion.⁵⁰

B. FORMAL INDEPENDENCE, INFORMAL ENTANGLEMENT

Notwithstanding platforms’ status as private actors, government preferences continue to shape firms’ internal content moderation systems, rules, and practices in a more general sense. Yet these kinds of pressures rarely amount to the kind of government coercion extensive enough to amount to a plausible First Amendment claim.⁵¹ It can be difficult to draw a line between changes to content-related decisions that occur because of jawboning and those that occur because of reputational or business risk.⁵² Using the examples of terrorist content and sex work, this subpart shows that firms’ behavior might be attributed as much to political climate as to unambiguous legal obligations.

1. Terrorist Content

Platforms have touted their ability to use artificial intelligence, automation, and hash matching to detect and prevent the dissemination of online terrorist content, advertising their abilities to proactively remove ISIS and al-Qaeda

47. Josh Gerstein, *Activists Troubled by White House Call to YouTube*, POLITICO (Sept. 14, 2012, 4:42), <https://www.politico.com/blogs/under-the-radar/2012/09/activists-troubled-by-white-house-call-to-youtube-135618>.

48. *Id.*

49. *Id.*

50. However, as I have previously argued elsewhere, threats of regulation can also generate “voluntary” proactive measures by platforms. Bloch-Wehba, *supra* note 10, at 58.

51. See Montgomery, *supra* note 35, at 172–73.

52. Kreimer, *supra* note 14, at 50; Ass’n of Am. Physicians & Surgeons v. Schiff, CV 20-106 (RC), 2021 WL 354174, at *6 (D.D.C. Feb. 2, 2021) (concluding that plaintiffs could not demonstrate that congressional statements led to private action by social media companies that lessened traffic to plaintiffs’ website).

terrorist content.⁵³ Notwithstanding claims of technical sophistication, however, critics have observed that platforms continued to allow designated foreign terrorist organizations such as Hamas, Hezbollah, and the FARC to maintain profiles and post content online, long after becoming aware of their activities.⁵⁴ Still, the numerous attempts to hold social media companies liable for the proliferation of online terrorist content have been unsuccessful.⁵⁵

From one perspective, in the absence of liability, social media firms have allowed themselves to be used as conduits for terrorist speech.⁵⁶ At the same time, however, firms have continued to engage in what Alexander Tsesis calls “corporate self-policing.”⁵⁷ For example, Zoom cancelled a 2020 San Francisco State University event with Leila Khaled, a Palestinian activist and member of the Popular Front for the Liberation of Palestine, a designated foreign terrorist organization.⁵⁸ Zoom argued that providing the platform for the talk would have violated federal laws prohibiting providing material support to terrorist organizations.⁵⁹ Similarly, in 2021, Google reportedly terminated the account of an activist sharing materials regarding Palestine on Google Drive, also citing violations of terrorism laws.⁶⁰

In many instances, these decisions go above and beyond what the law appears to require. In numerous cases, courts have held that platforms are not civilly liable when their services are used by terrorists.⁶¹ While regulators have pressured platforms to take more proactive steps to address terrorist content, platforms’ blunt approaches to removing terrorist content also risk over

53. Bloch-Wehba, *supra* note 10, at 59.

54. Citron & Wittes, *supra* note 26, at 403; Luis Jaime Acosta, *Social Networks Clamp Down on Colombian FARC Dissident Accounts*, REUTERS (Jan. 15, 2021, 11:40 AM), <https://www.reuters.com/N/us-twitter-colombia-idUSKBN29K2HI>.

55. *See, e.g.*, Fields v. Twitter, 881 F.3d 739 (9th Cir. 2018); Gonzalez v. Google, 282 F. Supp. 3d 1150 (N.D. Cal. 2017); Crosby v. Twitter, 303 F. Supp. 3d 564 (E.D. Mich. 2018); Force v. Facebook, 934 F.3d 53 (2d Cir. 2019).

56. Alexander Tsesis, *Social Media Accountability for Terrorist Propaganda*, 86 FORDHAM L. REV. 605, 611 (2017–18).

57. *Id.* at 613.

58. Alice Speri & Sam Biddle, *Zoom Censorship of Palestine Seminars Sparks Fight Over Academic Freedom*, THE INTERCEPT (Nov. 14, 2020, 4:00 AM), <https://theintercept.com/2020/11/14/zoom-censorship-leila-khaled-palestine/>.

59. *Id.*

60. @hotgirlhala, TWITTER (Apr. 22, 2021, 2:41 PM), <https://twitter.com/hotgirlhala/status/1385212069679702020>.

61. *See, e.g.*, Fields v. Twitter, 881 F.3d 739 (9th Cir. 2018); Crosby v. Twitter, 303 F. Supp. 3d 564 (E.D. Mich. 2018), *aff’d* 921 F.3d 617 (6th Cir. 2019); Force v. Facebook, 934 F.3d 53 (2d Cir. 2019); Sinclair for Tucker v. Twitter, Inc., C 17-5710 SBA, 2019 WL 10252752 (N.D. Cal. Mar. 20, 2019); Clayborn v. Twitter, Inc., 17-CV-06894-LB, 2018 WL 6839754 (N.D. Cal. Dec. 31, 2018).

ensorship.⁶² At the same time, this focus on Islamic terrorism, and particularly on ISIS and al-Qaeda, led to a severely under inclusive approach to other threats. Like many law enforcement agencies, social media companies paid little attention to problems of White nationalism and White extremism until after the Christchurch attacks in 2019.⁶³ Even then, platforms' mechanisms to address White nationalism and White supremacy have been haphazard and incomplete.⁶⁴

There are several potential explanations for platforms' voluntary actions to address terrorism. Most notable, perhaps, is the adoption of new regulations in Europe that require platforms to take down terrorist content within an hour, or else face liability.⁶⁵ In the United States, other pressures are in play. The threat that social media companies may face potential criminal liability under the material support statutes, as Tsesis and others have urged, may have encouraged platforms to address terrorism more aggressively.⁶⁶ Or perhaps the burgeoning calls to rethink Section 230's liability shield have led platforms to be more proactive, even in the absence of regulatory change. But government interests also provide powerful motivation for businesses to address harmful online content even when firms face no legal obligation to do so. Firms'

62. See ELEC. FRONTIER FOUN., SYRIAN ARCHIVE & WITNESS, *Caught in the Net: The Impact of Extremist Speech Regulations on Human Rights Content* (2019), <https://syrianarchive.org/en/lost-found/impact-extremist-human-rights#content-moderation-and-extremist-content> (last visited Aug. 9, 2021) [hereinafter *Caught in the Net*] (describing how reliance on automated tools to block and delete "terrorist content" also suppress human rights reporting, journalism, and other socially valuable posts).

63. Bloch-Wehba, *supra* note 10, at 60; Amna Akbar, *Policing Radicalization*, 3 UC IRVINE L. REV. 809, 827 (2013) (describing how indicators of Muslim religious observance were transmuted into signals of "radicalization").

64. See, e.g., Alex Kaplan, *YouTube Removed Some Channels Affiliated with White Nationalism—But Not All*, MEDIA MATTERS FOR AMERICA, <https://www.mediamatters.org/white-nationalism/youtube-removed-some-channels-affiliated-white-nationalism-not-all> (last visited June 22, 2021); Julia Carrie Wong, *White Nationalists are Openly Operating on Facebook. The company Won't Act*, THE GUARDIAN (Nov 21, 2019, 11:00 GMT), <http://www.theguardian.com/technology/2019/nov/21/facebook-white-nationalists-ban-vdare-red-ice>.

65. Regulation 2021/784 of the European Parliament and of the Council of Apr. 29, 2021, On Addressing The Dissemination of Terrorist Content Online ("TERREG"), annex, 2021 O.J. (L 172). In prior work, I have explored how platforms reacted to the emergence of new obligations in Europe, which have since been codified in the TERREG. See Bloch-Wehba, *supra* note 8, at 43–48 (detailing the evolution of European rules and platform responses on terrorist content).

66. Tsesis, *supra* note 56, at 625–26 (arguing that the material-support statute could support charges against recalcitrant social media service providers); Benjamin Wittes & Zoe Bedell, *Tweeting Terrorists, Part I: Don't Look Now but a Lot of Terrorist Groups are Using Twitter*, LAWFARE (Feb. 14, 2016, 5:05 PM), <https://www.lawfareblog.com/tweeting-terrorists-part-i-dont-look-now-lot-terrorist-groups-are-using-twitter>.

takedown priorities appeared to align with the government's law enforcement interests: in the context of a now decades-long war on (Islamic) terror, platforms likewise prioritized takedowns of ISIS and al-Qaeda content.⁶⁷

2. Sex Work

The experience of adult service businesses offers another illustration. In 2013, the Department of Justice initiated what it called “Operation Choke Point,” a program meant to encourage financial institutions to take a more active role in curtailing fraudulent businesses’ access to the banking system.⁶⁸ Critics of the program soon began to worry that banks were also cutting off legitimate businesses that they simply found distasteful, like pornographers, gun dealers, and payday lenders.⁶⁹ Although an audit later found that the Federal Deposit Insurance Corporation (FDIC) had not wrongly pressured banks to drop “high-risk” clients, it acknowledged that the agency’s regulatory activities “created a perception among some bank executives. . . that the FDIC discouraged institutions” from pursuing or maintaining business relationships with high-risk merchants.⁷⁰ In 2017, the Trump administration announced that it would put a stop to Operation Choke Point.⁷¹

Even in the absence of any legal requirements, many banks and payment processors have chosen to avoid providing services to adult businesses, perhaps because of social pressure or perceptions of other business risks.⁷² However, although online payment processors have no legal obligation to deny service to “high-risk” adult services clients, they nevertheless continue to keep

67. Bloch-Wehba, *supra* note 10, at 76–77; Caroline Mala Corbin, *Terrorists are Always Muslim but Never White: At the Intersection of Critical Race Theory and Propaganda*, 86 *FORDHAM L. REV.* 455, 458–60 (2017) (describing how popular culture, media narratives, and government priorities link “terrorism” to Muslim identity).

68. Richard P. Eckman, Richard J. Zack, Christina O. Hud, Jonathan N. Ledsky & Scott J. Helfand, *Update on the Short-Term Lending Industry: Government Investigations and Enforcement Actions*, 70 *BUS. LAW.* 657 (2014–2015).

69. Elizabeth Nolan Brown, *DOJ’s ‘Operation Choke Point’ may be Root of Porn Star Bank Account Closings*, *REASON.COM* (Apr 29, 2014, 8:40 PM), <https://reason.com/2014/04/28/doj-operation-chokepoint-and-porn-stars/>.

70. FED. DEPOSIT INS. CORP. OFF. OF INSPECTOR GEN., *The FDIC’s Role in Operation Choke Point and Supervisory Approach to Institutions that Conducted Business with Merchants Associated with High-Risk Activities*, at 11 (2015), <https://www.fdicoinc.gov/publications/fdics-role-operation-choke-point-and-supervisory-approach-institutions-conducted> (last visited Dec. 29, 2021).

71. Victoria Guida, *Justice Department to End Obama-era ‘Operation Choke Point’*, *POLITICO* (Aug. 17, 2017, 10:41 PM), <https://politi.co/2lObBHh>.

72. See E. Christopher Johnson, Jr., *The Important Role for Socially Responsible Businesses in the Fight Against Human Trafficking and Child Labor in Supply Chains*, *BUSINESS LAW TODAY* (Jan. 22, 2015), https://www.americanbar.org/groups/business_law/publications/blt/2015/01/02_johnson/.

adult services providers at arm's length, reflecting the perception of legal or business risk caused by providing such services to “high-risk” clients.⁷³ And payment processors are powerful intermediaries, critical to “people’s practical ability to speak”—and in this case, to post photos and videos or to maintain an online presence at all.⁷⁴

New intermediary obligations have heightened the sense that businesses must do more to address adult content. In 2018, Congress enacted the Allow States and Victims to Fight Online Sex Trafficking Act (FOSTA), a statute designed to promote platform accountability for sex trafficking.⁷⁵ As Eric Goldman has written, FOSTA responded to an apparent accountability gap that had allowed Backpage, a website primarily used for commercial sex advertising, to profit from advertisements of trafficking victims.⁷⁶ FOSTA expanded federal criminal liability for sex trafficking and for intentionally promoting or facilitating prostitution through interactive computer services.⁷⁷ Yet recent reporting suggests that FOSTA has hardly changed prosecutors’ ability to charge and convict sex traffickers. A 2021 Government Accountability Office report indicates that, in the past three years, prosecutors had only brought one case under FOSTA’s criminal provision.⁷⁸ In addition, as of June 2021, civil damages have never been awarded under FOSTA.⁷⁹

Despite its apparently sparse impact on criminal and civil liability, FOSTA clearly discouraged online platforms from hosting sexual content. In the wake of FOSTA’s passage, as Goldman recounts, several online service providers determined that they could no longer bear the risk of hosting *any* adult content at all. In 2018, online marketplace Craigslist stopped hosting personal ads entirely, citing the risk of criminal liability under FOSTA if adult content was

73. Sarah Manavis, *The PayPal ASMR banning Shows Us that Tech Companies Don't Understand Their Users*, NEWSTATESMAN (Sept. 20, 2018), <https://www.newstatesman.com/science-tech/technology/2018/09/paypal-asmr-ban-youtube-monetise-patreon> (describing broad application of PayPal’s sexual content policy); Margot Cleveland, *How Mastercard's Rules Could be Used to Ban Conservatives from Banking*, FEDERALIST (Apr. 19, 2021), <https://thefederalist.com/2021/04/19/how-mastercards-rules-against-child-pornographers-could-be-used-to-ban-conservatives-from-banking/>.

74. Jack M. Balkin, *Free Speech is a Triangle*, 118 COLUM. L. REV. 2011, 2014–15 (2018).

75. See Aja Romano, *A new law Intended to Curb Sex Trafficking Threatens the Future of the Internet as we Know It*, VOX (updated July 2, 2018, 1:08 PM EDT), <https://www.vox.com/culture/2018/4/13/17172762/fosta-sesta-backpage-230-internet-freedom>.

76. Eric Goldman, *The Complicated Story of Fosta and Section 230*, 17 FIRST AMEND. L. REV. 279, 281 (2018).

77. *Id.* at 284; 18 U.S.C. § 2421A.

78. *Sex Trafficking: Online Platforms and Federal Prosecutions* 25–26, Gov’t Accountability Office, GAO-21-385 (June 21, 2021), <https://www.gao.gov/products/gao-21-385>.

79. *Id.*

posted.⁸⁰ Even OnlyFans, the pay-per-view website known for risqué content, has a strict policy against escorts that has dramatically affected the livelihoods of sex workers.⁸¹ While OnlyFans hosts adult content by amateurs and celebrities, sex workers report being shunned by the platform, perhaps because of its assessment of the risk of liability under FOSTA.⁸² More broadly, sex workers have reported that FOSTA's enactment has "increased their exposure to violence and left those who rely on sex work as their primary form of income without many of the tools they had used to keep themselves safe."⁸³

Broadly speaking, then, both the examples of terrorist content and sex work illustrate that, even without an obvious enforcement mechanism, laws can encourage platforms to take aggressive private action against certain forms of speech, in alignment with government's own priorities. Some scholars might view this as a form of coercion or "jawboning," as Derek Bambauer and others have argued.⁸⁴ But others might describe platforms' actions here as the result of a more subtle form of government influence rather than a clear result of ham-fisted proxy censorship.⁸⁵ And when private incentives align with public policy, private governance provides a powerful new mechanism by which government can obtain its desired results without costly inconveniences such as accountability or oversight.

80. Merrit Kennedy, *Craigslist Shuts Down Personals Section After Congress Passes Bill on Trafficking*, NPR (Mar. 23, 2018, 3:52 PM), <https://www.npr.org/sections/thetwo-way/2018/03/23/596460672/craigslist-shuts-down-personals-section-after-congress-passes-bill-on-trafficking>; see also Heidi Tripp, *All Sex Workers Deserve Protection: How FOSTA/SESTA Overlooks Consensual Sex Workers in an Attempt to Protect Sex Trafficking Victims*, 124 PENN. ST. L. REV. 219 (2019) ("[M]any ISPs completely shut down certain services on their websites or began over-censoring content beyond what was necessary to comply with FOSTA/SESTA.").

81. Shae Ashbury, *How OnlyFans Steals from Sex Workers and Fans*, (Aug. 13, 2019), <https://www.shae-ashbury.com/shae-ashburys-blog/2019/8/13/how-onlyfans-steals-from-sex-workers>; Mark Serrels, *Thanks to US laws, Sex Workers are Fighting to Stay Online*, CNET (Feb. 26, 2021), <https://www.cnet.com/features/thanks-to-us-laws-sex-workers-are-fighting-to-stay-online/>.

82. Natalie Jarvey, *How OnlyFans Has Become Hollywood's Risque Pandemic Side Hustle*, HOLLYWOOD REP. (Dec. 11, 2020, 7:00 AM), <https://www.hollywoodreporter.com/news/how-onlyfans-has-become-hollywoods-risque-pandemic-side-hustle>; see also Alexis Okeowo, *The Fragile Existence of Sex Workers During the Pandemic*, NEW YORKER (May 21, 2021), <https://www.newyorker.com/news/news-desk/the-fragile-existence-of-sex-workers-during-the-pandemic> (describing how sex workers began to post on OnlyFans after SESTA-FOSTA); Serrels, *supra* note 81.

83. Danielle Blunt & Ariel Wolf, *Erased: The Impact of FOSTA-SESTA & the Removal of Backpage*, HACKING//HUSTLING 1 (2020), <https://hackinghustling.org/erased-the-impact-of-fosta-sesta-2020/>.

84. See Bambauer, *supra* note 37, at 891–99, 943.

85. See, e.g., Janice Nadler, *Expressive Law, Social Norms, and Social Groups*, 42 L. & SOC. INQUIRY 60, 64 (2017).

C. NEW INCENTIVES FOR PLATFORMS?

At one level, platforms have seemed eager to demonstrate their willingness and capacity to carry out government priorities through private policing. Yet governments (particularly outside of the United States) have also struggled to incentivize platforms to address unlawful content more aggressively. In the aftermath of the March 2019 massacre at two Christchurch mosques, governments proposed and adopted new legislation imposing penalties on online platforms that fail to remove unlawful content.⁸⁶

Both law enforcement and platforms see the potential for artificial intelligence and other automated techniques to enhance compliance with these measures and speed up takedowns. In Australia, for example, the law now imposes criminal penalties on providers of online services that do not remove “abhorrent violent material” “expeditiously.”⁸⁷ In Germany, the Network Enforcement Act of 2018 similarly requires platforms to quickly remove unlawful content, sometimes within 24 hours, or pay large fines.⁸⁸ The European Union recently finalized its regulation on terrorist content online, which will not only require platforms to take down terrorist content more quickly, but also require them to adopt more proactive measures to prevent the spread of terrorist content in the first place.⁸⁹

These kinds of pressures have led free speech advocates and scholars to see in government regulatory proposals the clear threat of proxy censorship. The dominant accounts of law enforcement interests in this space describe governments as seeking more extensive takedowns, more limits on speech, and more aggressive enforcement of private and public rules, while platforms resist the imposition of these and similar obligations.⁹⁰ Faced with platforms’ independence and immunity from liability, governments seek to require them to behave more aggressively in filtering out unlawful content, ideally through adopting new technologies of decision-making.

86. Evelyn Douek, *Australia’s New Social Media Law is a Mess*, LAWFARE (Apr. 10, 2019, 8:28 AM), <https://www.lawfareblog.com/australias-new-social-media-law-mess>.

87. Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019 (Cth) § 474.34 (Austl.).

88. Evelyn Douek, *Germany’s Bold Gambit to Prevent Online Hate Crimes and Fake News Takes Effect*, LAWFARE (Oct. 31, 2017, 11:30 AM), <https://www.lawfareblog.com/germanys-bold-gambit-prevent-online-hate-crimes-and-fake-news-takes-effect>.

89. Regulation 2021/784 of the European Parliament and of the Council of Apr. 29, 2021, On Addressing The Dissemination of Terrorist Content Online (“TERREG”), annex, 2021 O.J. (L 172).

90. See generally Evelyn Douek, *Australia’s “Abhorrent Violent Material” Law: Shouting “Nerd Harder” and Drowning Out Speech*, 94 AUSTL. L. REV. 41 (2020).

Yet powerful interests also cut in the opposite direction, encouraging platforms to keep unlawful content online as a form of intelligence for law enforcement to mine. The enactment of FOSTA has reportedly made it much more difficult for law enforcement to investigate and detect sex trafficking victims and perpetrators.⁹¹ Similarly, mechanisms for removing terrorist content have diminished the availability of human rights reporting online.⁹² By contrast, law enforcement has a strong interest in maintaining access to social media's trove of online evidence. The more aggressively social media platforms enforce their private rules, whether through automated technology or through manual review, the harder it becomes for law enforcement to conduct this kind of surveillance.⁹³

III. PLATFORMS' INFLUENCE ON POLICING

The communicative and data-generating affordances of online platforms change user behavior and create legal challenges.⁹⁴ In turn, they also drive investigative strategy.⁹⁵ In the previous Part, I demonstrated that law enforcement sometimes seeks to control or influence the affordances of social media platforms, especially when it comes to dangerous or violent speech. But as this Part shows, the content-related decision-making of platforms also benefits law enforcement, creating new sources of information with new affordances for investigating online speech. As a result, police increasingly depend upon purportedly private content moderation rules, strategies, and techniques, and platforms have a growing role in facilitating law enforcement surveillance. The aim here is to complicate what has become a binary distinction between platform and government and illustrate the mutual entanglements of the two.

91. See Appellant's Br. at 54, *Woodhull v. DOJ*, No. 18-5298 (D.C. Cir. filed Feb. 13, 2019); Masnick, *supra* note 13.

92. *Caught in the Net*, *supra* note 62.

93. See *infra* text accompanying notes 110–111.

94. For example, the ability to upload user-generated content to YouTube has facilitated widespread copyright infringement. García, *supra* note 23, at 285–86. Surprisingly, rightsholders have sometimes encouraged infringement because they benefit from the free publicity and “Internet buzz.” *Id.* at 298–99.

95. Cf. Orin S. Kerr, *An Equilibrium-Adjustment Theory of the Fourth Amendment*, 125 HARV. L. REV. 476, 482 (2011) (describing how “changing technology” and “social practice” might impel courts to respond by ratcheting up or down the rules that constrain police power).

A. SHAPING LAW ENFORCEMENT THROUGH TECHNOLOGY

It is widely appreciated that private governance plays an increasing role in state policy.⁹⁶ As Jack Balkin has observed, the growing capacity of internet firms to surveil and control content has also made them “more valuable targets” for regulation.⁹⁷ Perhaps less appreciated, however, is the degree to which the affordances of networked technologies increasingly shape law enforcement practices themselves. Design choices dictate what information is available to law enforcement—and thus what information law enforcement can demand and use in investigative contexts.⁹⁸ Law enforcement is engaged in a form of what Marion Fourcade and Jeffrey Gordon call “dataist statecraft,” in which the availability of data minted by both public and private actors drives policy.⁹⁹

1. Compelled Disclosure

Historically, many of the fights about law enforcement access to user information have been about compelled disclosure of customer records and communications.¹⁰⁰ The law of compelled disclosure governs the standards by which law enforcement can obtain access to different categories of user information in the possession of firms. For example, the Stored Communications Act (SCA) imposes a warrant requirement for communications that have been stored in an electronic communications system for 180 days or less.¹⁰¹ If communications have been stored for greater than 180 days, then the government can seek access using a subpoena, court order, or a search warrant, accompanied by different indicia of suspicion and different notice obligations.¹⁰² This legal structure has generated numerous

96. Balkin, *supra* note 74, at 2028 (“[N]ew-school speech regulation depends on the expansion and promulgation of private governance.”); Robert Gorwa, *The Platform Governance Triangle: Conceptualising the Informal Regulation of Online Content*, 8 INTERNET POL’Y REV. 1, 7 (2019) (“[I]nformal regulatory arrangements have formed a key tool through which governance stakeholders—especially EU governments—have sought to shape the behaviour of firms on content issues.”).

97. Balkin, *supra* note 74, at 2020.

98. Andrew Guthrie Ferguson, *Structural Sensor Surveillance*, 106 IOWA L. REV. 47, 54 (2020) (“[D]esign choices can directly impact the usefulness of the data collected.”).

99. Marion Fourcade & Jeffrey Gordon, *Learning Like a State: Statecraft in the Digital Age*, 1 J. L. & POL. ECON. 78, 78 (2020).

100. *See generally* Kerr, *supra* note 15, at 1209–12 (describing how ambiguities in the Fourth Amendment’s application to the Internet fostered legal uncertainty about compelled disclosure of user communications); *see also In re 381 Search Warrants Directed to Facebook, Inc.*, 29 N.Y.3d 231 (2017).

101. 18 U.S.C. § 2703(a).

102. 18 U.S.C. § 2703(b); *see also* Hannah Bloch-Wehba, *Exposing Secret Searches: A First Amendment Right of Access to Electronic Surveillance Orders*, 93 WASH. L. REV. 145, 160 (2018)

legal battles regarding whether disclosure of a range of information—including historic cell site location data, web browsing histories, and the contents of emails—implicates the Fourth Amendment’s definition of a search or seizure.¹⁰³

Focusing on the appropriate standard for defining a government search, though, threatens to miss the degree to which the availability of networked technologies itself drives law enforcement strategy. Digital searches and seizures have vastly grown in number, reflecting the increased relevance of digital communications in investigations, the growing scale of networked technology applications and services, and the proliferation of different forms of information.¹⁰⁴ As the number of requests for user information has increased, the role of electronic communications service providers in facilitating, obstructing, and enabling surveillance has also grown apace.¹⁰⁵ Providers such as Google, Facebook, and Microsoft have large in-house compliance teams in order to process a growing number of law enforcement requests for customer data.¹⁰⁶

Networked technologies are not only driving an increase in the degree of law enforcement surveillance and control; they are also fundamentally transforming the work of law enforcement. Consider the surveillance of cell site location information. In the last decade, significant ink has been spilled regarding law enforcement’s acquisition of cell phone location information through real-time tracking, historic location data, cell tower dumps, cell site simulators, and data purchases.¹⁰⁷ In 2018, the Supreme Court decided in

(describing the different notice provisions for different forms of legal process to compel disclosure of customer records and communications).

103. *Carpenter v. United States*, 138 S. Ct. 2206, 2217 (2018) (holding that a disclosure of a week of cell site location information is a search); *United States v. Forrester*, 512 F.3d 500, 511 (9th Cir. 2008) (holding that acquiring IP addresses of websites user visited is not a search); *United States v. Warshak*, 631 F.3d 266, 288 (6th Cir. 2010) (holding that a disclosure of email contents is a search).

104. Rozenshtein, *supra* note 15, at 109 (arguing that digital intermediaries are “more central than ever to government surveillance”).

105. *See id.* at 114; Manes, *supra* note 15, at 348.

106. *See, e.g., Facebook Transparency Report*, <https://transparency.facebook.com/government-data-requests/country/US> (last visited June 21, 2021) (documenting a rise in the number of requests from 11,000 in the first half of 2013 to over 61,000 in the first half of 2020); *Google Transparency Report*, <https://transparencyreport.google.com/user-data/overview> (last visited July 15, 2021) (documenting a rise in the number of requests from 25,000 in the first half of 2013 to over 100,000 in the first half of 2020).

107. Brian L. Owsley, *The Fourth Amendment Implications of the Government’s Use of Cell Tower Dumps in its Electronic Surveillance*, 16 UNIV. PA. J. CONST. L. 1 (2013); Stephen Wm. Smith, *Gagged, Sealed & Delivered: Reforming ECPA’s Secret Docket*, 6 HARV. L. & POL’Y REV. 601 (2012); Andrew Guthrie Ferguson, *The Internet of Things and the Fourth Amendment of Effects*, 104 CALIF.

Carpenter v. United States that obtaining over six days of historical cell site location from a cell phone service provider constitutes a search for Fourth Amendment purposes.¹⁰⁸ However, the *Carpenter* Court expressly declined to decide whether cell tower dumps, which collect all the phone numbers that connected to a given cell tower during a given time period, held the same Fourth Amendment implications.¹⁰⁹ Soon afterward, law enforcement began to seek so-called geofence or reverse location information—information pertaining to every user in a given geographical radius during a given time period—from Google.¹¹⁰ Google’s location tracking—infamously difficult to turn off or opt out of—becomes the new equivalent of the cell tower.¹¹¹ Networked technologies, by design, collect and retain information from large numbers of users, in turn driving law enforcement to seek more data from these sources.¹¹²

Likewise, the emergence of the so-called Internet of Things and omnipresent embedded sensors are equally responsible for novel transformations in investigative strategy.¹¹³ Law enforcement can now acquire data from connected speakers, fitness trackers, doorbell cameras, and smart streetlights.¹¹⁴ As a result, consumer technology may drive not only self-

L. REV. 805 (2016); Byron Tau, *House Investigating Company Selling Phone Location Data to Government Agencies*, WALL ST. J: POLITICS (June 24, 2020, 3:19 PM), <https://www.wsj.com/articles/house-investigating-company-selling-phone-location-data-to-government-agencies-11593026382>.

108. *Carpenter v. United States*, 138 S. Ct. 2206 (2018).

109. *Id.* at 2220; *see also* Owsley, *supra* note 107, at 16–17 (arguing that cell tower dumps are more intrusive than simple pen registers).

110. *See, e.g., In re Search Warrant Application for Geofence Location*, 497 F. Supp. 3d 345 (N.D. Ill. 2020).

111. Ryan Nakashima, *Google Tracks Your Movements, Like it or Not*, AP NEWS (Aug. 14, 2018), <https://apnews.com/article/north-america-science-technology-business-ap-top-news-828aefab64d4411bac257a07c1af0ecb>.

112. *See, e.g., Carpenter*, 138 S. Ct. at 2218 (“[T]he retrospective quality of the data here gives police access to a category of information otherwise unknowable.”).

113. Paul Ohm, *The Many Revolutions of Carpenter*, 32 HARV. J. L. & TECH. 357, 364–65 (2019).

114. Kayla Epstein, *Police Think Amazon’s Alexa may have Information on a Fatal Stabbing Case*, WASH. POST (Nov. 2, 2019), <https://www.washingtonpost.com/technology/2019/11/02/police-think-amazons-alexa-may-have-information-fatal-stabbing-case/>; Christine Hauser, *Police Use Fitbit Data to Charge 90-Year-Old Man in Stepdaughter’s Killing*, N. Y. TIMES (Oct. 3, 2018), <https://www.nytimes.com/2018/10/03/us/fitbit-murder-arrest.html>; John Herrman, *Who’s Watching Your Porch?*, N. Y. TIMES (Jan. 19, 2020), <https://www.nytimes.com/2020/01/19/style/ring-video-doorbell-home-security.html>; Jesse Marx, *Police Used Smart Streetlight Footage to Investigate Protesters*, VOICE SAN DIEGO (June 29, 2020), <https://www.voiceofsandiego.org/topics/government/police-used-smart-streetlight-footage-to-investigate-protesters/>.

tracking, but law enforcement tracking as well.¹¹⁵ As discussed in Part IV, networked, sensory technologies do not just create goldmines of information for law enforcement, but also fundamentally alter the legal mechanisms through which policing can be made transparent and accountable to the public.

2. “Open Source” Investigations

The growing role of compelled disclosure in law enforcement investigations illustrates the centrality of networked technology as a mechanism of surveillance, but it is just the tip of the iceberg. Although law enforcement can influence platform rules and practices through either takedown requirements (as in Part II) or compelled disclosure requirements (as in Part III.A), social media can also influence law enforcement by serving as a ready source of open-source information and evidence.

For instance, law enforcement regularly monitors public social media activity in both targeted investigations and as a source of dragnet intelligence.¹¹⁶ As advocates at the Brennan Center have explained, social media surveillance often occurs when officers “view[] publicly available posts by searching for an individual, group, hashtag, or another search vector.”¹¹⁷ The extent, scope, and manner in which these results might be displayed depends on the affordances of the platform at issue. For example, the U.S. Department of Homeland Security has monitored Black Lives Matter groups and events using Twitter hashtags and location information.¹¹⁸

Law enforcement has also used surveillance services such as Geofeedia, Snaprends, and others to access social media data in an automated fashion.¹¹⁹

115. GINA NEFF & DAWN NAFUS, SELF-TRACKING 178 (Mass. Inst. Tech., 2016) (describing the potential legal questions around self-tracking data).

116. Rachel Levinson-Waldman, *Private Eyes, They're Watching You: Law Enforcement's Monitoring of Social Media*, 71 OKLA. L. REV. 997, 999–1000 (2019); Rachel Levinson-Waldman, *Government Access to and Manipulation of Social Media: Legal and Policy Challenges*, 61 HOW. L.J. 523, 541–42 (2018); see also Daphna Renan, *The Fourth Amendment as Administrative Governance*, 68 STAN. L. REV. 1039, 1053 (2016) (noting that “generalized collection” can lead to targeted surveillance).

117. Rachel Levinson-Waldman & Ángel Díaz, *How to Reform Police Monitoring of Social Media*, BROOKINGS (July 9, 2020), <https://www.brookings.edu/techstream/how-to-reform-police-monitoring-of-social-media/>.

118. George Joseph, *Exclusive: Feds Regularly Monitored Black Lives Matter Since Ferguson*, THE INTERCEPT (July 24, 2015, 11:50 AM), <https://theintercept.com/2015/07/24/documents-show-department-homeland-security-monitoring-black-lives-matter-since-ferguson/>.

119. Matt Cagle, *Facebook, Instagram, and Twitter Provided Data Access for a Surveillance Product Marketed to Target Activists of Color*, ACLU: N. CAL. (Oct. 11, 2016), <https://www.aclunc.org/blog/facebook-instagram-and-twitter-provided-data-access-surveillance-product-marketed-target>; see also Ryan Devereaux, *Homeland Security Used a Private Intelligence Firm to Monitor Family Separation Protests*, THE INTERCEPT (Apr. 19, 2019, 8:25 AM), <https://theintercept.com/2019/>

Social media surveillance can allow law enforcement agencies to assess social media information for potential risks and threats and map connections between investigative targets and other subjects.¹²⁰ On the other hand, police do not always recognize the gravity of online threats or “chatter.” On January 5, 2021, Dataminr reached out to police at the U.S. Capitol to notify them of an uptick in chatter regarding the upcoming riots, but law enforcement reportedly took no preparatory action.¹²¹

To some extent, the emergence of third-party social media surveillance tools like Dataminr and Geofeedia is a direct response to platform firms’ user affordances and content policies. In April 2021, the New York Police Department (NYPD) published a draft “impact and use policy” for public comment on NYPD’s social media surveillance systems.¹²² The policy stressed that NYPD only accesses “publicly available information, or information that is viewable as a result of user privacy settings or practices.” However, the policy also explained that third party surveillance tools help to fill critical investigative gaps that result when users or platforms delete content relevant to an investigation.¹²³

04/29/family-separation-protests-surveillance/; Colin Daileida, *Twitter Cuts Ties with Another Social Media Surveillance Company*, MASHABLE, (Oct. 20, 2016) <https://mashable.com/article/twitter-social-media-surveillance-snaptrands>; Colin Daileida, *Geofeedia isn’t the Only Social Media Surveillance Company Giving Data to Police*, MASHABLE, (Oct. 12, 2016) <https://mashable.com/article/geofeedia-social-media-surveillance-police>.

120. JOHN HOLLYWOOD, MICHAEL JOHN DEVRIES VERMEER, DULANI WOODS, SEAN GOODISON & BRIAN JACKSON, USING SOCIAL MEDIA AND SOCIAL NETWORK ANALYSIS IN LAW ENFORCEMENT: CREATING A RESEARCH AGENDA, INCLUDING BUSINESS CASES, PROTECTIONS, AND TECHNOLOGY NEEDS 8–9 (Rand Corporation, 2018) (describing social media monitoring for “worrisome activity” and in order to identify individuals “at high risk of being involved in violence”).

121. Zachary Cohen & Whitney Wild, *Internal Emails Reveal Capitol Security Officials Dismissed Warnings About Troubling Social Media Posts Before January 6 Riot*, CNN (Apr. 28, 2021, 6:16 AM), <https://www.cnn.com/2021/04/28/politics/capitol-security-emails-social-media-riot/index.html>.

122. Police Department City Of New York, *Social Network Analysis Tools: Impact and Use Policy* (Apr. 11, 2021), https://www1.nyc.gov/assets/nypd/downloads/pdf/public_information/post-final/social-network-analysis-tools-nypd-impact-and-use-policy_4.9.21_final.pdf. The policy was published pursuant to the Public Oversight of Surveillance Technology Act, a law enforcement reform bill that requires the New York Police Department to publish reports about its “surveillance technology.”; Hannah Bloch-Wehba, *Visible Policing: Technology, Transparency, and Democratic Control*, 109 CALIF. L. REV. 917, 955 (2021).

123. POLICE DEP’T N.Y.C., *Social Network Analysis Tools: Impact and Use Policy* 3 (Apr. 11, 2021), https://www1.nyc.gov/assets/nypd/downloads/pdf/public_information/post-final/social-network-analysis-tools-nypd-impact-and-use-policy_4.9.21_final.pdf. (“NYPD may miss information critical to investigations because users can easily remove information posted on social media and social media platforms routinely delete content and deactivate accounts for violations of terms of service. Accordingly, social network analysis tools allow the NYPD

In addition to dragnet surveillance of events and people of interest, law enforcement uses social media in more targeted ways, often in contexts in which police rely upon undercover operations and confidential informants.¹²⁴ For instance, local police and the Federal Bureau of Investigation (FBI) have reportedly used undercover social media accounts to surveil groups and individuals and to develop probable cause to arrest suspected lawbreakers.¹²⁵ Similarly, law enforcement routinely uses fake social media accounts to engage in investigations.¹²⁶ For example, sex trafficking investigators often create fake social media accounts to “befriend, identify, and monitor people suspected of engaging in criminal activities, as well as those who are presumed to be victims.”¹²⁷ Although creating a fake social media account often violates a platform’s terms of service and other content-related rules, this practice appears prevalent.

3. Deputizing Users

The public-facing character of social media itself can feed into law enforcement strategies. Although law enforcement frequently monitors social media content and demands access to the wealth of data that online firms collect and retain, police also engage with users much as other ordinary users

to retain information on social networking platforms relevant to investigations and alert investigators to new activity on queried social media accounts.”).

124. See Cyrus Farivar & Olivia Solon, *FBI Trawled Facebook to Arrest Protestors for Inciting Riots, Court Records Show*, NBC NEWS (June 19, 2020, 1:26 PM), <https://www.nbcnews.com/tech/social-media/federal-agents-monitored-facebook-arrest-protesters-inciting-riots-court-records-n1231531> (describing FBI’s use of social media to “infiltrate activist groups”).

125. Betsy Woodruff Swan, *Feds Comb Facebook to Hunt down Alleged Rioters and Looters*, POLITICO (June 12, 2020, 4:30 AM), <https://www.politico.com/news/2020/06/12/facebook-riot-loot-justice-department-314567>.

126. See, e.g., Dave Maass, *Facebook Warns Memphis Police: No More Fake “Bob Smith” Accounts*, ELECTR. FRONTIER FOUND. (Sept. 24, 2018), <https://www.eff.org/deeplinks/2018/09/facebook-warns-memphis-police-no-more-fake-bob-smith-accounts>; Dave Maass, *Four Steps Facebook Should Take to Counter Police Sock Puppets*, ELECTR. FRONTIER FOUND. (Apr. 14, 2019), <https://www.eff.org/deeplinks/2019/04/facebook-must-take-these-four-steps-counter-police-sock-puppets>; Jon Schuppe, *Undercover Cops Break Facebook Rules to Track Protestors, Ensnare Criminals*, NBC NEWS, (Oct. 5, 2018, 12:08 PM), <https://www.nbcnews.com/news/us-news/undercover-cops-break-facebook-rules-track-protesters-ensnare-criminals-n916796>; Tami Abdollah, *U.S. Plan to Use Fake Social Media Profiles for Surveillance is Against Facebook Rules*, PBS (Sept. 3, 2019, 5:19 PM), <https://www.pbs.org/newshour/nation/u-s-plan-to-use-fake-social-media-profiles-for-surveillance-is-against-facebook-rules>.

127. JENNIFER MUSTO, CONTROL AND PROTECT: COLLABORATION, CARCERAL PROTECTION, AND DOMESTIC SEX TRAFFICKING IN THE UNITED STATES 57–58 (1st ed. 2016).

do.¹²⁸ And, of course, law enforcement uses social media to disseminate routine information to a mass audience.¹²⁹

Law enforcement also relies on social media to generate tips and investigative leads. For instance, police sometimes post videos to social media to solicit the public's help in identifying a suspect.¹³⁰ This form of crowdsourced public assistance can be crucial for investigating crimes but raises complex questions about online vigilantism, anonymity, and accountability. In the wake of the January 6 putsch at the U.S. Capitol, the FBI called for "the public's assistance in identifying individuals who made unlawful entry into the U.S. Capitol building and committed various other alleged criminal violations."¹³¹ Though the vast majority of the insurrectionists walked away from the scene at the Capitol, social media users, private investigators, and the press identified dozens of individuals who were later charged.¹³² This is not the first time in which a group of self-appointed internet users have tried to identify and hold accountable lawbreakers. In 2017, after the Unite the Right Rally in Charlottesville, online sleuths identified and outed, or "doxxed," several right-wing and White supremacist protestors.¹³³ But online vigilantes sometimes identify the wrong people, leading to harassment of innocent

128. See Levinson-Waldman, *Private Eyes*, *supra* note 116, at 999 ("[I]f a targeted user has a public Twitter account, police can go on the site to check the user's recent posts and interactions with other users without needing any special third-party software.").

129. Benesch, *supra* note 9, at 93 ("The very functions of routine governance are also carried out, increasingly, on social media platforms."); see also Knight First Amendment Inst. at Colum. Univ. v. Trump, 928 F.3d 226, 235–36 (2019) (describing how President Trump used his Twitter account "as an important tool of governance and executive outreach").

130. *Aggravated Assault 1 South Broad St. Dc 21 06 015773*, YOUTUBE (May 6, 2021), <https://www.youtube.com/watch?v=FCeCu8WT3es>; *Severe Injury Hit and Run Traffic Collision in Northeast Area NR21122nw*, YOUTUBE (May 5, 2021), <https://www.youtube.com/watch?v=zNoDjrw5W-M>.

131. FBI, *U.S. Capitol Violence*, <https://www.fbi.gov/wanted/capitol-violence> (last visited June 21, 2021); see also FBI Washington Field (@FBIWFO), TWITTER, <https://twitter.com/FBIWFO/status/1347407275300954112> (last visited Dec. 29, 2021).

132. Jaclyn Peiser, *Internet Detectives are Identifying Scores of Pro-Trump Rioters at the Capitol. Some have Already been Fired.*, WASH. POST (Jan. 8, 2021, 6:54 AM), <https://www.washingtonpost.com/nation/2021/01/08/capitol-rioters-fired-doxed-online/>; Sara Morrison, *The Capitol Rioters Put Themselves All Over Social Media. Now they're Getting Arrested.*, VOX, <https://www.vox.com/recode/22218963/capitol-photos-legal-charges-fbi-police-facebook-twitter> (last updated: Jan 19, 2021, 6:52 PM); Greg Myre, *How Online Sleuths Identified Rioters at the Capitol*, NPR (Jan. 11, 2021, 9:45 AM) <https://www.npr.org/2021/01/11/955513539/how-online-sleuths-identified-rioters-at-the-capitol>.

133. Vegas Tenold, *To Doxx a Racist: How a Dead White Supremacist Sparked the Debate About the Tactics Used Against the Extreme Right*, THE NEW REPUBLIC (July 26, 2018), <https://newrepublic.com/article/150159/doxx-racist>; Emma Grey Ellis, *Whatever Your Side, Doxing is a Perilous Form of Justice—Even When it's Outing Nazis*, WIRED (Aug. 17, 2017, 8:00 AM), <https://www.wired.com/story/doxing-charlottesville/>.

individuals. For example, in 2013, users of the subreddit Find Boston Bombers misidentified several people as suspects in the Boston Marathon attack.¹³⁴

Scholars such as Mary Anne Franks and Danielle Citron have warned that doxing can be part of a campaign of online harassment and abuse, with particularly devastating results for women.¹³⁵ But unlike these earlier episodes, the hundreds of Capitol putsch arrests and prosecutions appear to rely heavily on identifications made using information gleaned from social media, whether crowdsourced or obtained directly from platforms.¹³⁶ Indeed, in light of major social media platforms' decisions to take down much of the evidence related to the Capitol putsch, crowdsourcing may have been particularly essential to identifying individuals.¹³⁷

The public also uses social media to alert law enforcement to suspicious activity in more mundane settings. Consider Nextdoor, a social media platform designed for “neighbors” to exchange information with each other.¹³⁸ It is a unique surveillance tool because it facilitates voluntary, private surveillance by those who choose to join the platform.¹³⁹ As Sam Levin has documented, White Nextdoor users have deployed the platform to report unsubstantiated claims of suspicious activity and to organize noise complaints against Black

134. Dave Lee, *Boston Bombing: How Internet Detectives Got it Very Wrong*, BBC NEWS: TECHNOLOGY (Apr. 19, 2013), <https://www.bbc.com/news/technology-22214511>.

135. Mary Anne Franks, *Sexual Harassment 2.0*, MD. L. REV. 655, 678–79 (2012) (describing episode of sexual harassment in which online forum users “posted personal information of their targets” and encouraged forum participants to contact victims directly); DANIELLE KEATS CITRON, HATE CRIMES IN CYBERSPACE 53–54 (2016); *see also* David M. Douglas, *Doxing: A Conceptual Analysis*, 18 ETHICS INFO. TECH. 199, 200 (2016) (“In cases where exposing wrongdoing is in the public interest, deanonymizing and delegitimizing doxing is permissible only to the extent necessary to reveal that wrongdoing has occurred.”).

136. Craig Timberg, Drew Harwell & Spencer S. Hsu, *Police Let Most Capitol Rioters Walk Away. But Cellphone Data and Videos Could Now Lead to More Arrests.*, WASH. POST (Jan. 8, 2021), <https://www.washingtonpost.com/technology/2021/01/08/trump-mob-tech-arrests/> (“The countless hours of video—much of it taken by the rioters themselves and uploaded to social media—also offers an ideal data set for facial recognition.”); Drew Harwell & Craig Timberg, *How America’s Surveillance Networks Helped the FBI Catch the Capitol Mob*, WASH. POST (Apr. 2, 2021), <https://www.washingtonpost.com/technology/2021/04/02/capitol-siege-arrests-technology-fbi-privacy/>.

137. Knibbs, *supra* note 12 (describing how efforts to preserve online documentation of the Capitol putsch related to FBI’s efforts to seek evidence for use in criminal proceedings); *How Facebook is Responding to the Violence at the US Capitol*, FACEBOOK (Jan. 11, 2021, 1:00 PM), <https://www.facebook.com/business/news/facebooks-actions-in-response-to-washington-dc-violence> (describing Facebook’s actions to remove content that “incites, praises, or encourages violence or harm,” including support for the Capitol putsch).

138. *About Nextdoor*, <https://about.nextdoor.com/> (last visited June 21, 2021).

139. Rahim Kurwa, *Building the Digitally Gated Community: The Case of Nextdoor*, 17 SURVEILLANCE & SOC’Y 111, 113 (2019) (describing the “co-production of community through participation in surveillance”).

residents.¹⁴⁰ In response to criticism that its platform was amplifying patterns of racial profiling and harassment, Nextdoor adopted changes to its service to discourage users from posting unsubstantiated, racialized accusations.¹⁴¹ In 2020, during nationwide uprisings against police violence, Nextdoor announced that it was removing its “Forward to Police” feature, which permitted users to forward posts directly to law enforcement partners.¹⁴² Although Nextdoor has tried to nudge users away from using crime reporting to perpetuate racial harassment, crime prevention is still a core part of Nextdoor’s offerings and appeal.¹⁴³

4. Resistance Through Design

While this Article focuses on how platforms can enable and facilitate law enforcement surveillance, in recent years, firms have also made design choices that can obstruct policing, generating substantial legal and political backlash. Although these choices can take many forms, I highlight two here.

First, firms can choose to collect and store data about user communications in ways that are more or less vulnerable to law enforcement demands.¹⁴⁴ For example, Signal, a secure messaging provider, simply “does

140. Sam Levin, *Racial Profiling via Nextdoor.Com*, EAST BAY EXPRESS (Oct. 7, 2015), <https://eastbayexpress.com/racial-profiling-via-nextdoorcom-2-1/>.

141. Sam Levin, *What Happens when Tech Firms End Up at the Center of Racism Scandals?*, THE GUARDIAN (Aug. 30, 2016), <http://www.theguardian.com/technology/2016/aug/30/tech-companies-racial-discrimination-nextdoor-airbnb> (describing Nextdoor’s adoption of a new system that warns users about racial profiling before they post a crime and safety message); see also Tatyana Mamut, *Announcing Our New Feature to Promote Kindness in Neighborhoods*, NEXTDOOR: BLOG (Sept. 18, 2019), <https://blog.nextdoor.com/2019/09/18/announcing-our-new-feature-to-promote-kindness-in-neighborhoods/> (describing Nextdoor’s “Kindness Reminder” feature, which nudges users to reconsider offensive or hurtful posts before publishing); Team Nextdoor, *Standing in Solidarity with Black Neighbors—Nextdoor*, (Mar. 25, 2021), <https://blog.nextdoor.com/2021/03/25/standing-in-solidarity-with-black-neighbors/> (prohibiting All Lives Matter and Blue Lives Matter content “when used to undermine racial equality or the Black Lives Matter movement”).

142. Team Nextdoor, *Nextdoor Removes “Forward to Police” Feature*, NEXTDOOR: BLOG (June 18, 2020), <https://blog.nextdoor.com/2020/06/18/nextdoor-removes-forward-to-police-feature/>.

143. Joseph Porcelli, *Nextdoor for Public Agencies Crime Prevention Engagement Plan*, MEDIUM (May 22, 2019), <https://medium.com/nextdooragencyresources/nextdoor-for-public-agencies-crime-prevention-engagement-plan-1bf92c34b360>; see, e.g., Timothy Hayden, Arlington Policy Department, *Requesting Assistance in Identifying Suspect that Broke into Vehicles in Your Neighborhood*, (Apr. 19, 2021), <https://nextdoor.com/agency-post/tx/arlington/arlington-police-department/requesting-assistance-in-identifying-suspect-that-broke-into-vehicles-in-your-neighborhood-184037858/>.

144. SHOSHANNA ZUBOFF, SURVEILLANCE CAPITALISM 385 (2019) (describing how government officials “must work, at least in part, through the [private] surveillance capitalists” to access and make use of consumer data).

Encryption has become a major point of contention for law enforcement in the United States and elsewhere. Domestically, for example, the San Bernardino shootings generated legal controversy when Apple refused to unlock the shooter's iPhone.¹⁵² In 2020, several Republican senators introduced the Lawful Access to Encrypted Data (LAED) Act, which “would bring an end to warrant-proof encryption in devices, platforms, and systems.”¹⁵³ The LAED Act would “require device manufacturers and service providers to assist law enforcement with accessing encrypted data if assistance would aid in the execution of the warrant.”¹⁵⁴ Similar approaches have been adopted elsewhere. In the United Kingdom, the Investigatory Powers Act permits the government to issue a “technical capability notice” that requires firms to be able to assist in executing lawful warrants.¹⁵⁵

As outlined above, technological design choices like these have prompted substantial controversy and strife between regulators and platforms. Both encryption and data storage choices can make it more difficult for platforms or law enforcement to access information about user speech that is either harmful or unlawful.¹⁵⁶ However, some automated mechanisms for screening

3A2420600258234172%7D&path=%2Fnotes%2Fnote%2F&refsrc=http%3A%2F%2Ft.co%2F&_rdr (last visited June 22, 2021) (announcing plans to work on end-to-end encryption); *but see* Andy Greenberg, *Facebook Says Encrypting Messenger by Default Will Take Years*, WIRED (Jan. 10, 2020, 4:54 PM), <https://www.wired.com/story/facebook-messenger-end-to-end-encryption-default/>.

152. Ellen Nakashima & Reed Albergotti, *The FBI Wanted to Unlock the San Bernardino Shooter's iPhone. It Turned to a Little-Known Australian Firm.*, WASH. POST (Apr. 14, 2021), <https://www.washingtonpost.com/technology/2021/04/14/azimuth-san-bernardino-apple-iphone-fbi/>.

153. United States Senate Committee on the Judiciary, *Graham, Cotton, Blackburn Introduce Balanced Solution to Bolster National Security, End Use of Warrant-Proof Encryption that Shields Criminal Activity*, <https://www.judiciary.senate.gov/press/rep/releases/graham-cotton-blackburn-introduce-balanced-solution-to-bolster-national-security-end-use-of-warrant-proof-encryption-that-shields-criminal-activity> (last visited Dec. 29, 2021).

154. *Id.* The EARN IT Act proposed in 2020 adopted a similar approach, requiring platforms to qualify for a statutory safe harbor under Section 230 of the Communications Decency Act by showing that they abided by “best practices” to fight child sexual exploitation. As several commentators noted, those “best practices” were likely incompatible with strong encryption.; Lily Hay Newman, *The EARN IT Act is a Sneak Attack on Encryption*, WIRED (Mar. 5, 2020, 8:22 PM), <https://www.wired.com/story/earn-it-act-sneak-attack-on-encryption/>; Riana Pfefferkorn, *The EARN IT ACT is a disaster amid the COVID-19 crisis*, BROOKINGS INST. (May 4 2020), <https://www.brookings.edu/techstream/the-earn-it-act-is-a-disaster-amid-the-covid-19-crisis/>.

155. Investigatory Powers Act 2016 § 253.

156. Michael H. Keller & Gabriel J. X. Dance, *The Internet is Overrun with Images of Child Sexual Abuse. What Went Wrong?*, N.Y. TIMES (Sept. 28, 2019), <https://www.nytimes.com/interactive/2019/09/28/us/child-sex-abuse.html> (“[W]hen tech companies cooperate fully, encryption and anonymization can create digital hiding places for perpetrators.”).

content may be compatible with end-to-end encryption.¹⁵⁷ Whether firms choose to deploy them is an entirely different design question.

B. SHAPING LAW ENFORCEMENT THROUGH PLATFORM POLICY

Like technological design, firms' internal policies also shape law enforcement behavior by encouraging or discouraging certain kinds of demands for different types of data. Partly because of the First Amendment implications of compelled disclosure, social media platforms have sometimes resisted government demands, citing the implications for their users. For example, technology companies have, at times, moved to quash government search warrants, attempting to advance the Fourth Amendment interests of their users.¹⁵⁸ Electronic communications service providers have also invoked their own expressive rights in efforts to lift nondisclosure orders that prevent service providers from notifying users of demands for users' information.¹⁵⁹

Outside of litigation, firms can also engage in private standard-setting to raise the standard that the government must meet when it demands user information.¹⁶⁰ Again, consider the example of government requests for historical location information. As Matthew Tokson has observed, the Supreme Court's *Carpenter* decision is "exceedingly vague and cautious" with regard to its application to new technologies and forms of surveillance.¹⁶¹ Therefore, substantial ambiguity remains about whether government activity constitutes a search or a seizure.¹⁶²

157. Jonathan Mayer, *Content Moderation for End-to-End Encrypted Messaging*, 5 (Oct. 6, 2019), https://www.cs.princeton.edu/~jrmayer/papers/Content_Moderation_for_End-to-End_Encrypted_Messaging.pdf.

158. *See, e.g., In re* 381 Search Warrants Directed to Facebook, Inc., 29 N.Y.3d 231 (2017).

159. *See, e.g., Microsoft Corp. v. United States Dep't of Justice*, 233 F. Supp. 3d 887, 908 (W.D. Wash. 2017) (concluding that Microsoft had adequately supported its argument that nondisclosure orders under the Electronic Communications Privacy Act violated its First Amendment rights); *In re Nat'l Sec. Letter*, 863 F.3d 1110 (9th Cir. 2017) (rejecting petitioner's First Amendment challenge to nondisclosure orders that accompanied National Security Letters); *Twitter, Inc. v. Barr*, 445 F. Supp. 3d 295 (N.D. Cal. 2020) (rejecting Twitter's First Amendment challenge to the government's prohibition on publishing certain types of data regarding legal process the platform had received under the Foreign Intelligence Surveillance Act).

160. *Cf. Klonick, supra* note 5, at 1615 (developing the idea of private governance in the context of platform content moderation standards).

161. Matthew Tokson, *The Next Wave of Fourth Amendment Challenges after Carpenter*, 59 WASHBURN L.J. 1, 1 (2020).

162. *Id.; see also United States v. Hammond*, 996 F.3d 374, 391–92 (7th Cir. 2021) (concluding that real-time collection of location information for several hours was not a "search" in the meaning of the Fourth Amendment).

In the wake of *Carpenter*, does the government's demand for location information from Google raise a Fourth Amendment issue? At times, private platform policymaking can preempt this inquiry. In one case, police investigating a string of fires sought information from Google regarding user devices near six different locations.¹⁶³ Although the government has argued that *Carpenter* does not extend to reverse location information, in practice, Google will only provide this information in response to a search warrant.¹⁶⁴ That creates a default practice in which the government must satisfy a higher modicum of suspicion to satisfy Google's policy notwithstanding the absence of controlling legal precedent.

Firms also use the mechanisms of private governance—particularly policies and terms of service—to limit government access to data in other ways. In 2016, the American Civil Liberties Union (ACLU) obtained information, through public records requests, showing that law enforcement agencies were procuring social media monitoring software from third-party vendors. In response, major social media firms such as Facebook, Twitter, and Instagram publicly announced that they would cut off access to their application programming interfaces (APIs) by firms that sold surveillance software to law enforcement.¹⁶⁵

Yet it appears that these policy-based limitations on government access are often ineffective. During the nationwide uprisings against police violence after George Floyd's murder in 2020, for example, it became clear that law enforcement agencies were continuing to use social media monitoring services such as Dataminr to keep tabs on protest, activism, and dissent.¹⁶⁶ In order to circumvent Twitter's terms of service, which barred API users from "tracking, alerting, or monitoring sensitive events," Dataminr has rebranded itself as a breaking news service that takes advantage of access to Twitter's "firehose" to provide alerts to law enforcement clients.¹⁶⁷ The upshot is that while platform firms' formal content and privacy policies appear to bar use of their services

163. *In re* Search Warrant Application for Geofence Location, 497 F. Supp. 3d at 351.

164. *See In re* Search of Info. Stored at Premises Controlled by Google, 481 F. Supp. 3d 730, 736 (N.D. Ill. 2020) (finding that, because the government had sought a search warrant, it had "forfeited the argument" that the Fourth Amendment didn't apply); *see also id.* *In re* Search Warrant Application for Geofence Location, 497 F. Supp. 3d at 360 (noting that Google "will only produce the information upon presentation of a warrant").

165. Levinson-Waldman, *Government Access*, *supra* note 116, at 556–57.

166. Biddle, *supra* note 11; *see also* Lee Fang, *FBI Expands Ability to Collect Cellphone Location Data, Monitor Social Media, Recent Contracts Show*, THE INTERCEPT (June 24, 2020, 8:56 PM), <https://theintercept.com/2020/06/24/fbi-surveillance-social-media-cellphone-dataminr-venntel/>; Sahar F. Aziz & Khaled A. Beydoun, *Fear of a Black and Brown Internet: Policing Online Activism*, 100 B.U. L. REV. 1151, 116869 (2020).

167. Biddle, *supra* note 11.

for law enforcement surveillance, a veritable cottage industry of surveillance and monitoring firms has sprung up to help police take full advantage of the wealth of intelligence that social media can provide.

C. VOLUNTARY PRIVATE-PUBLIC SURVEILLANCE ARRANGEMENTS

It is not just that private governance sometimes serves as an ineffective check on law enforcement; at times, private decision-making can in fact advance law enforcement goals. Indeed, voluntary private decision-making can give rise to a systematic relationship with law enforcement investigations, arrests, and prosecutions. Faced with competing pressures to both take down more harmful content and to facilitate law enforcement surveillance, the private sector has increasingly turned to voluntary, cross-platform arrangements that allow them to pool technical and policy resources across firms.¹⁶⁸

Collaboration to eradicate child sexual abuse imagery provides one illustration. While technology firms are not required to proactively monitor user-uploaded or -generated content for unlawful child sexual abuse imagery, many do so voluntarily.¹⁶⁹ For example, Microsoft's PhotoDNA program, a hash-matching tool, scans images and videos against a database of unlawful images.¹⁷⁰ Thorn, a nonprofit organization, has developed a technical tool for the same purposes for smaller companies to use.¹⁷¹ When a firm detects a match, federal law requires the firm to report it to the National Center for Missing and Exploited Children (NCMEC).¹⁷² NCMEC, in turn, discloses the information to law enforcement and plays an essential coordinating role with law enforcement agencies investigating the crime.¹⁷³ While NCMEC itself is a private organization, it is funded through annual grants by the government and, pursuant to federal law, must coordinate several distinct public and private programs.¹⁷⁴ At least one federal court has concluded that NCMEC's

168. EVELYN DOUEK, *THE RISE OF CONTENT CARTELS* 5–6 (Knight First Amendment Inst. Colum. Univ., 2020) (describing voluntary cross-industry arrangements as “content cartels”).

169. Bloch-Wehba, *supra* note 10, at 58.

170. *Id.* at 58.

171. Olivia Solon, *To Fight Online Child Sexual Abuse, Tech Companies Turn to a Nonprofit Startup*, NBC NEWS (July 22, 2020, 3:16 PM), <https://www.nbcnews.com/tech/tech-news/fight-online-child-sexual-abuse-tech-companies-turn-nonprofit-startup-n1234569>.

172. 18 U.S.C. § 2258A.

173. *United States v. Ackerman*, 831 F.3d 1292, 1296 (10th Cir. 2016).

174. 34 U.S.C. § 11293(b).

statutory obligations give rise to “special law enforcement duties and powers” that distinguish it from other private entities.¹⁷⁵

Platforms also collaborate on efforts to filter and block terrorist content online. Consider the Global Internet Forum to Counter Terrorism (GIFCT), a private, voluntary consortium of technology firms that uses both hash-based and artificial intelligence-based filtering to detect unlawful terrorist content across platforms.¹⁷⁶ The GIFCT’s database is limited to industry members and is not shared directly with law enforcement.¹⁷⁷ But new European regulations will require platforms to take proactive measures to remove terrorist content and to preserve it for law enforcement purposes for six months.¹⁷⁸ The result is that the GIFCT hash-matching database is likely to yield a substantial number of posts that platforms will be required to preserve for potential law enforcement use and possibly to report to the “competent authorities.”¹⁷⁹

These two examples illustrate a strikingly similar dynamic: technology firms have voluntarily adopted monitoring technology to enforce content-related rules, the use of which gives rise to an escalating set of legal obligations. In the context of child sexual abuse imagery, platforms must report information to NCMEC, a nominally private center, which then funnels it to law enforcement.¹⁸⁰ In the context of terrorist imagery, platforms are required to report certain kinds of terroristic threats to European authorities and likewise required to preserve a broader range of information for future law enforcement use.¹⁸¹ The result is that the monitoring technology used to detect

175. See *Ackerman*, 831 F.3d at 1296–97. A second court has concluded that NCMEC can act as part of the “prosecution team” for purposes of discovery, and of obligations to disclose exculpatory evidence pursuant to *Brady v. Maryland*; *United States v. Rosenschein*, CR 16-4571 JCH, 2019 WL 2298810, at *7 (D.N.M. May 30, 2019), clarified on denial of reconsideration, CR 16-4571 JCH, 2020 WL 2750247 (D.N.M. May 27, 2020) (“It was NCMEC’s acts of investigating the location and providing CyberTipline information to the geographically appropriate law enforcement agency that effectively commenced the prosecution of this case.”).

176. *Explainers*, GLOBAL INTERNET F. COUNTER TERRORISM, <https://gifct.org/explainers/> (last visited June 21, 2021); Robert Gorwa, Reuben Binns & Christian Katzenbach, *Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance*: 7 BIG DATA & SOCIETY 1 (2020) (describing GIFCT’s use of hash-based and machine learning techniques).

177. GLOB. INTERNET F. COUNTER TERRORISM, *supra* note 176.

178. Regulation 2021/784 of the European Parliament and of the Council of Apr. 29, 2021, On Addressing The Dissemination of Terrorist Content Online (“TERREG”), art. 6 sec. 3, art. 3.

179. *Id.* art. 6(1)–(2) (requiring hosting service providers to “preserve terrorist content” for six months); art. 14(5).

180. *Ackerman*, 831 F.3d at 1296–97.

181. TERREG arts. 6 and 14.

lawbreaking itself lies at the heart of investigations and prosecutions, yielding increasing entanglements between law enforcement and platform governance.¹⁸²

The increased reliance on automated content moderation also creates new opportunities and frameworks within which platforms share user communications with law enforcement. Ordinarily, the Stored Communications Act (SCA) bars electronic communications service providers from voluntarily disclosing user communications to law enforcement, with a few exceptions.¹⁸³ Broadly speaking, the SCA is meant to limit the circumstances in which user communications are shared with law enforcement to those circumstances in which the government has met the appropriate standard.¹⁸⁴ But if the communications service provider obtains the contents of communications “inadvertently” and they “appear to pertain to the commission of a crime,” then the provider may disclose the contents to a law enforcement agency.¹⁸⁵

Firms that detect legal violations using technical moderation tools are arguably free to voluntarily disclose that information to law enforcement pursuant to the SCA because they learned of it “inadvertently” and the contents “appear to pertain to the commission of a crime.”¹⁸⁶ Alternatively, law enforcement could use a search warrant, administrative subpoena, or 2703(d) order to compel a platform to disclose subscriber information for any user who has uploaded content that has been flagged as unlawful.¹⁸⁷ Law enforcement has pursued this dragnet approach before. In 2017, the government obtained a search warrant to compel DisruptJ20, a website that had been used to organize protests against Donald Trump’s inauguration, to disclose records related to a huge number of people who had visited the site.¹⁸⁸ It is well within the realm of possibility that law enforcement may use a similar process to seek information about individuals who have been flagged through

182. *See, e.g.*, *State v. Lizotte*, 197 A.3d 362, 366 (Vt. 2018) (describing AOL’s use of its “Image Detection Filtering Process” in the context of a defense motion to suppress).

183. 18 U.S.C. § 2702(b).

184. *See* 18 U.S.C. § 2701(a)–(c) (making it a criminal offense to access stored communications, except if doing so is authorized); 18 U.S.C. § 2703(b)–(c) (setting forth procedural requirements that law enforcement must meet in order to access different types of communications information).

185. 18 U.S.C. § 2702(b)(7)(A).

186. *Id.*

187. 18 U.S.C. § 2703.

188. Coalition: Justice Department’s demand for protest website data raises privacy and civil liberty concerns, *OPENTHEGOVERNMENT.ORG* (Aug. 24, 2017), <https://www.openthegovernment.org/coalition-justice-departments-demand-for-protest-website-data-raises-privacy-and-civil-liberty-concerns/>.

the GIFCT database or who have been suspended for posting terrorist-related content.

IV. IMPLICATIONS FOR CRIMINAL PROCEDURE

As Reidenberg rightly anticipated, today's public sphere is shaped as much by private technological and design choices as by formal law and regulation.¹⁸⁹ But the emergence of private platforms as regulatory forces in their own right has not uniformly diminished the role or power of the state. Certainly, platform intransigence on content-related issues has, at times, posed challenges for law enforcement.¹⁹⁰ But platforms can also expand and facilitate law enforcement power by encoding and enforcing law enforcement demands in the rules, norms, and technological infrastructures of online governance.

The current alignment between private technology firms and public law enforcement has expanded the authority and the power of both firms and states. At the same time, as governments seek to incentivize technology firms to prevent the dissemination of unlawful speech through private governance and technology, they also enlist technology firms to aid in digital surveillance, both directly and indirectly.¹⁹¹ As outlined in Parts II and III, these efforts can sometimes occur at cross-purposes; increasing deletion of online content has, at times, created obstacles for law enforcement investigating criminal activity online. But here, too, technology itself can provide a workaround. New surveillance technology tools and practices emerge, taking advantage of the affordances of social media for law enforcement's gain.

A. THE EMERGENCE OF NEW FORMS OF DISCLOSURE

New legal and technological developments only underscore the mutual dependency between private firms and the public sector. As both governments and tech firms herald the growing capacity and use of artificial intelligence and automated content moderation systems, content- and data-related decision-making itself is increasingly becoming entwined with law enforcement objectives. Yet the more extensive public-private cooperation becomes, the weaker the opportunities for accountability appear.

189. Reidenberg, *supra* note 16, at 571 ("The political-governance process ordinarily establishes the substantive law of the land. For Lex Informatica, however, the primary source of default rule-making is the technology developer and the social process by which customary uses evolve.").

190. See, e.g., Danielle Keats Citron & Benjamin Wittes, *The Problem isn't Just Backpage: Revising Section 230 Immunity*, 2 GEO. L. TECH. REV. 453, 466 (2018) (listing myriad bad actors who were protected from legal liability under Section 230 of the Communications Decency Act).

191. Balkin, *supra* note 74, at 2019–20.

Emerging forms of content and data governance generate new demands by law enforcement for consumer data.¹⁹² Again, the example of the Internet of Things is illustrative: why conduct a search of a person's home in real time when a set of networked home technologies makes it possible to do so retrospectively? As online platforms turn increasingly toward automated moderation techniques to proactively filter user-generated content, the content moderation process itself becomes an increasingly appealing target for law enforcement. Platforms engaged in automated content moderation will obtain access to a huge amount of content that either violates or appears to violate the law. Indeed, the explicit goal of automated moderation is to scale enforcement of platform rules and practices to respond to the growing volume of online content.¹⁹³ But because the technology is not yet that sophisticated, automated techniques are often necessarily over inclusive.¹⁹⁴ This means that, at times, automated moderation will sweep in more content than it was intended to.

In the United States, the examples of NCMEC and Thorn already illustrate how voluntary content moderation processes can feed law enforcement demands.¹⁹⁵ But existing laws governing the sharing of data between private and public sector actors are ill-equipped to address these emerging practices. The Stored Communications Act (SCA) presumptively limits the sharing of private user information between communications firms and law enforcement to a defined set of circumstances governed by appropriate statutory limitations. For example, the SCA explicitly provides that platforms may voluntarily disclose user data to NCMEC in connection with a statutorily required report regarding child sexual abuse imagery.¹⁹⁶ As tech firms engage in more extensive collaboration, including with nonprofits and independent organizations such as Thorn and GIFCT, neither the SCA nor the Fourth Amendment are likely to promote private accountability. For its part, the SCA's voluntary disclosure limitations extend only to actors who provide a "remote computing service" or "electronic communication service" *to the public*, which coalitions such as Thorn and GIFCT do not.¹⁹⁷ Moreover, the SCA explicitly permits platforms to share data among themselves without constraint.¹⁹⁸ The lax attitude toward private data sharing will encourage more voluntary, private arrangements to

192. See *infra* Part III.

193. Gorwa et al., *supra* note 176, at 2.

194. Gorwa et al., *supra* note 176, at 5 (describing how machine learning systems "risk over-blocking in cases in which the word may be acceptable in context").

195. See Ohm, *supra* note 113.

196. 18 U.S.C. § 2702(c)(5).

197. 18 U.S.C. § 2702(a).

198. 18 U.S.C. § 2702(c)(6).

emerge, while ignoring how those arrangements feed law enforcement demands for data.¹⁹⁹

Outside the United States, new statutory initiatives already make clear that law enforcement has a growing appetite to deputize the content moderation process in service of investigative needs. In Germany, legislators introduced a new version of the Network Enforcement Act (NetzDG) alongside a package of measures intended to strengthen criminal law enforcement.²⁰⁰ The new initiatives would require social network providers to report content that violated certain criminal prohibitions directly to law enforcement, along with the user's IP address and passwords.²⁰¹ As platforms continue to ramp up their efforts to police harmful and unlawful content through technology and through policy, the data they collect will, itself, become a rich source of evidence for law enforcement.

B. NEW INVESTIGATIVE METHODS

Firms' private decisions regarding both design and policy do not only shape law enforcement practices. They also shape the law of criminal procedure itself.

First, the design of technical infrastructure that facilitates the simultaneous collection and retention of information from large numbers of users encourages a shift from individualized suspicion to larger scale "dragnets," often with unclear consequences for Fourth Amendment protections.²⁰² For

199. Cf. Hannah Bloch-Wehba, *Transparency after Carpenter*, 59 WASHBURN L.J. 23, 28 (2020) ("[P]rivate sector data collection has created a rich source of information for law enforcement, yet goes hand in hand with stringent limitations on government conduct.").

200. Evelyn Douek, *Germany's Bold Gambit to Prevent Online Hate Crimes and Fake News Takes Effect*, LAWFARE (Oct. 31, 2017, 11:30 AM), <https://www.lawfareblog.com/germanys-bold-gambit-prevent-online-hate-crimes-and-fake-news-takes-effect>; Amelie Heldt, *Germany is Amending its Online Speech Act NetzDG. . . But Not Only That*, INTERNET POL'Y REV. (Apr. 6, 2020), <https://policyreview.info/articles/news/germany-amending-its-online-speech-act-netzdg-not-only/1464>.

201. Patrick Beuth, *Was Sie über das Gesetz gegen Hasskriminalität Wissen Müssen*, DER SPIEGEL, (Feb. 18, 2020, 5:10 PM) <https://www.spiegel.de/netzwelt/netzpolitik/gesetz-gegen-hasskriminalitaet-was-sie-darueber-wissen-muessen-a-1f995e2b-80a9-4e11-aecc-75f3250c69b9> (reporting that social network providers would be required to report certain violent threats, neo-Nazi propaganda, and incitement of hatred along with the IP addresses and port numbers of the subscribers to the German federal criminal police; in addition, providers may also be required to share passwords with law enforcement or intelligence agencies).

202. Renan, *supra* note 116, at 1053; Barry Friedman & Cynthia Benin Stein, *Redefining what's Reasonable: The Protections for Policing*, 84 GEO. WASH. L. REV. 281, 303–04 (2016) (describing the turn toward "dragnet searches"); Christopher Slobogin, *Government Dragnets*, 73 LAW & CONTEMP. PROBS. 107, 110 (2010) (defining "dragnets" as "programmatic government efforts to investigate, detect, deter, or prevent crime or other significant harm by subjecting a

example, information gleaned from generalized social media surveillance might be included in law enforcement databases and in targeted investigations. Gang policing is illustrative: Police frequently use social media information in gang databases, which collect and maintain information about alleged gang members.²⁰³ Posts in which a user “admits” to gang membership, photos that include gang signs or other alleged gang members, and “liking” other users’ gang related posts are all reportedly sufficient to land a social media user in a gang database.²⁰⁴ Data from social media also makes its way into predictive policing tools, immigration enforcement, and domestic terrorism investigations.²⁰⁵

Law enforcement often describes the scraping, analysis, and use of huge amounts of publicly available data to predict and control behavior as an essential tool for high-priority investigations.²⁰⁶ But perhaps this is exactly backwards—perhaps it is the availability of the data itself, and the possibilities for analysis and interpretation, that drive law enforcement’s turn toward new priorities.²⁰⁷ And, of course, it is not just the ease of acquiring privately held data that incentivizes law enforcement to adopt new data-driven techniques,

group of people, most of whom are concededly innocent of wrongdoing or of plans to engage in it, to a deprivation of liberty or other significant intrusion.”)

203. Hannah Bloch-Wehba, *Visible Policing: Technology, Transparency, and Democratic Control*, 109 CALIF. L. REV. 917, 950–54 (2021).

204. Sara Robinson, *When a Facebook Like Lands You in Jail*, BRENNAN CTR. FOR JUST. (July 6, 2018), <https://www.brennancenter.org/our-work/analysis-opinion/when-facebook-lands-you-jail>; STUART, *supra* note 39, at 9; Meredith Broussard, *When Cops Check Facebook*, THE ATLANTIC (Apr. 19, 2015), <https://www.theatlantic.com/politics/archive/2015/04/when-cops-check-facebook/390882/>.

205. Drew Harwell & Nick Miroff, *ICE Just Abandoned its Dream of “Extreme Vetting” Software that Could Predict whether a Foreign Visitor would become a Terrorist*, WASH. POST (May 17, 2018), <https://www.washingtonpost.com/news/the-switch/wp/2018/05/17/ice-just-abandoned-its-dream-of-extreme-vetting-software-that-could-predict-whether-a-foreign-visitor-would-become-a-terrorist/>; Will Carless, *Feds are Tracking Americans’ Social Media to Identify Dangerous Conspiracies. Critics Worry for Civil Liberties.*, USA TODAY (2021), <https://www.usatoday.com/story/news/nation/2021/05/14/terrorist-social-media-narratives-focus-new-dhs-effort/5075237001/>.

206. *See, e.g.*, Wes Simmons, *Big Data Does Not Have to Mean Big Brother or be a Big Deal*, POLICE CHIEF MAGAZINE (May 3, 2017), <https://www.policechiefmagazine.org/big-data-does-not-have-to-mean-big-brother/> (recounting hypothetical case of an illegally armed individual identified and prevented from committing a violent act because of big data); *see also* Andrew Guthrie Ferguson, *Big Data and Predictive Reasonable Suspicion*, 163 U. PA. L. REV. 327, 362–63 (2015) (describing how “law enforcement and private companies have embraced the idea of networking and sharing personal information”).

207. *Cf.* Fourcade & Gordon, *supra* note 99, at 79–80 (describing how technology generates “new possibilities,” often controlled and marketed to government by private firms).

but also the emergence of privately developed networked technologies that are themselves purpose-built for law enforcement uses.²⁰⁸

Second, the essential role of networked technology firms in facilitating surveillance also limits opportunities for oversight agencies, the public, and defendants to understand how law enforcement is doing its job. Once upon a time, the search of a home was the “canonical fact pattern” of Fourth Amendment law.²⁰⁹ But today, rather than conducting a physical search of one’s home, law enforcement can issue a remote request to a Silicon Valley firm to compel disclosure of reams of intimate data from inside the same four walls.²¹⁰ As Jon Michaels has pointed out with regard to intelligence, public-private cooperation can enhance secrecy and impede oversight.²¹¹ The web of sealing and secrecy orders that often surrounds electronic surveillance tends to obscure the role that social media platforms play in facilitating law enforcement investigations.²¹² In prior work, I have also suggested that the emergence of more secretive forms of surveillance has diminished the opportunities for defendants to use the law of criminal procedure to hold law enforcement accountable.²¹³

In theory, networked technology firms should reduce, rather than increase, secrecy. As David Pozen has put it, a secret is “deep” if its existence is concealed from the public; a secret is “shallow” if “ordinary citizens understand they are being denied relevant information and have some ability to estimate its content.”²¹⁴ Private firms, which exist outside the executive branch, are arguably able to increase and facilitate public knowledge of surveillance practices.²¹⁵

208. Elizabeth E. Joh, *The New Surveillance Discretion: Automated Suspicion, Big Data, and Policing*, 10 HARV. L. & POL’Y REV. 15, 15 (2016) (“New technologies have altered surveillance discretion by lowering its costs and increasing the capabilities of the police to identify suspicious persons.”); Andrew Guthrie Ferguson, *The Exclusionary Rule in the Age of Blue Data*, 72 VAND. L. REV. 561, 597 (2019).

209. Orin S. Kerr, *Searches and Seizures in a Digital World*, 119 HARV. L. REV. 531, 536 (2005).

210. Ohm, *supra* note 113, at 395–96 (arguing that *Carpenter* requires law enforcement to get a warrant before obtaining smart home data from third-party technology providers).

211. See Jon D. Michaels, *All the President’s Spies: Private-Public Intelligence Partnerships in the War on Terror*, 96 CALIF. L. REV. 901, 922–26 (2008) (describing how informal public-private partnerships can minimize leaks, negative publicity, and legal risk).

212. Smith, *supra* note 107, at 602; Manes, *supra* note 15, at 351.

213. Bloch-Wehba, *supra* note 122, at 14 (“Diminished Fourth Amendment protections have also made it much more difficult for courts, defendants, and the public to get critical information necessary to check the police.”).

214. David E. Pozen, *Deep Secrecy*, 62 STAN. L. REV. 257, 274 (2009–10).

215. Manes, *supra* note 15, at 344 (“If these companies could win the right to speak about the *kinds* of records the government is ordering them to disclose, they would be able to provide

Yet the increasing role of private sector actors in policing has not appreciably diminished law enforcement secrecy, but rather shifted the locus of claims of secrecy to private sector actors. Today, technology companies routinely invoke trade secrecy and other corporate protections to avoid transparency about the role they play in facilitating law enforcement investigations and prosecutions.²¹⁶ In an atmosphere of increased calls to defund and reform policing, secrecy plays an essential role in protecting law enforcement's private partners from reputational risk.²¹⁷ By shielding private firms from the reputational costs of partnering with police, however, law enforcement also reduces the public's ability to monitor and understand how the government conducts surveillance.

Third, technology firms are unconstrained by constitutional limitations. For Fourth Amendment purposes, constitutional constraints on searches and seizures are limited to “unreasonable searches undertaken by the government or its agents—not private parties.”²¹⁸ As Kiel Brennan-Marquez has documented, the traditional approach has been to examine whether a private party was “deputized” by the state to investigate; if so, the party loses its “private” status, and the Fourth Amendment applies to the search.²¹⁹ Brennan-Marquez points out, however, that courts have uniformly held that *voluntary* private hashing such as that accomplished by PhotoDNA software or the GIFCT is not covered by the Fourth Amendment.²²⁰ This creates a legal gap, inviting law enforcement to exploit private action and informal relationships to extend its own power. In contrast, Brennan-Marquez suggests that, where the government relies on private action to extend the infrastructure of surveillance, the Constitution ought to follow.²²¹

Here, regulatory action and jawboning have created new incentives for platforms to engage in private policing that itself drives law enforcement

the public with crucial information about how the surveillance laws have been interpreted and applied in practice.”).

216. See generally Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343 (2018); see also Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 87 (2019) (describing how Northpointe, a firm that provides risk assessment tools used at sentencing, conceals the weight of its risk scores based on trade secrecy); Natalie Ram, *Innovating Criminal Justice*, 112 NW. U. L. REV. 659, 668–71 (2018) (describing how, pursuant to non-disclosure agreements with Harris Corp., police concealed the use of Stingray technology).

217. Michaels, *supra* note 211, at 926 (arguing that “handshake collaborations” with government agencies may generate litigation risk for firms).

218. *Ackerman*, 831 F.3d at 1295.

219. Brennan-Marquez, *supra* note 17, at 488.

220. *Id.* at 504.

221. *Id.* at 505.

action. Yet Fourth Amendment protections have not followed. Voluntary private searches for unlawful content trigger no Fourth Amendment protection. Neither, seemingly, does the use of new technologies of surveillance that reimagine online speech as a source of law enforcement intelligence.²²² Indeed, the lack of constitutional constraints likely encourages informal relationships between tech firms and law enforcement by avoiding the heavy costs of the warrant requirement, reasonableness limitations, and the exclusionary rule.²²³ As Brennan-Marquez suggests, these circumstances may warrant a reimagining of the constitutional status of purportedly “private” searches.²²⁴

To be clear, I do not suggest that any time a platform takes action on content in a manner favored by the government, the Constitution ought to attach. Technology firms are powerful, wealthy actors; the fact that their interests sometimes (or even often) align with the government’s is hardly surprising, nor is it cause for inherent suspicion. But the extensive alignment between platforms and states ought to prompt scholars and policymakers to reexamine the prevalent assumption that technology firms are engaged in forms of private governance accountable to nobody except, possibly, their shareholders. The truth is that, while platforms can provide a powerful counterweight to state action, they can just as easily buttress it.

C. DESIGN AND LEGAL IMMUNITY

Another way in which the government might limit private accountability for partnering with law enforcement is through granting immunity from suit for firms. The government might simply decide to require technology firms to design their services and products in ways that are more amenable to law enforcement. For example, the Communications Assistance for Law Enforcement Act (CALEA) requires telecommunications carriers to design their services to be capable of providing access for law enforcement.²²⁵ As Gus

222. See Jeramie D. Scott, *Social Media and Government Surveillance: The Case for Better Privacy Protections for Our Newest Public Space*, 12 J. BUS. & TECH. L. 151, 158–59 (2017) (explaining that Fourth Amendment does not protect social media posts “knowingly expose[d]” to the public eye); cf. Ken Dilanian, *DHS Launches Warning System to Find Domestic Terrorism Threats on Public Social Media*, NBC NEWS (May 10, 2021, 10:30 AM), <https://www.nbcnews.com/politics/national-security/dhs-launches-warning-system-find-domestic-terrorism-threats-public-social-n1266707> (describing a Department of Homeland Security proposal to use social media to detect domestic terrorism threats).

223. Cf. William J. Stuntz, *Distribution of Fourth Amendment Privacy*, 67 GEO. WASH. L. REV. 1265, 1274–77 (1999) (observing that Fourth Amendment law makes certain investigative activities more “costly” than others).

224. Brennan-Marquez, *supra* note 17, at 489.

225. 47 U.S.C. § 1002(a).

Hurwitz has pointed out, CALEA was “arguably the first time that Congress had imposed affirmative design requirements on firms in order to support law enforcement capabilities.”²²⁶ Outside the United States, regulators are already pressuring platforms to redesign their moderation techniques and rules to prioritize law enforcement, as the NetzDG and the EU Terrorist Regulation both demonstrate. As outlined above, a growing chorus of proposals would also require platforms to retain and disclose user data for law enforcement as well.

Congress may also grant statutory immunity to firms that partner with law enforcement. Consider the response after the New York Times published a story in 2005 detailing how telecommunications companies had partnered willingly with federal law enforcement and intelligence agencies after September 11th to collect the contents of communications under what was known inside the Bush Administration as the “Terrorist Surveillance Program,” which became colloquially known as “warrantless wiretapping.”²²⁷ In the following months and years, dozens of lawsuits were filed against the telecommunications companies themselves, as well as the National Security Agency (NSA).²²⁸ In response, Congress enacted the Federal Intelligence Surveillance Act of 1978 Amendments Act of 2008 (“FISA Amendments Act”), which codified provisions authorizing the bulk collection of some foreign communications.²²⁹ The FISA Amendments Act also granted conditional statutory immunity to private firms that worked with the Terrorist Surveillance Program under assurances that the program was lawful.²³⁰ News organizations have also reported that telecommunications firms have assisted law enforcement with wiretapping in legally dubious circumstances after

226. Justin Hurwitz, *Encryption[^]Congress Mod(Apple + CALEA)*, 30 HARV. J. L. & TECH. 355, 379 (2017).

227. James Risen & Eric Lichtblau, *Bush Lets U.S. Spy on Callers without Courts*, N. Y. TIMES (Dec. 16, 2005), <https://www.nytimes.com/2005/12/16/politics/bush-lets-us-spy-on-callers-without-courts.html>.

228. *See, e.g., In re Nat'l Sec. Agency Telecomm. Rec. Litig.*, 671 F.3d 881, 890 (2011) (analyzing statutory immunity for telecommunications companies that collaborated with NSA warrantless wiretapping); *Jewel v. Nat'l Sec. Agency*, 673 F.3d 902, 906 (9th Cir. 2011) (permitting AT&T subscriber to proceed in First and Fourth Amendment challenge to warrantless wiretapping that AT&T conducted “in collaboration with the [NSA]”); *American Civil Liberties Union v. Nat'l Sec. Agency*, 493 F.3d 644 (6th Cir. 2007) (finding that plaintiffs lacked standing for First and Fourth Amendment claims against the NSA regarding the Terrorist Surveillance Program).

229. Pub. L. No. 110-261.

230. 50 U.S.C. § 1885a(4)(A); *In re Nat'l Sec. Agency Telecomm. Rec. Litig.*, 633 F. Supp. 2d 949, 959 (2009) (describing the FISA Amendments Act immunity provision as “sui generis” because of its limitations on subject-matter, time period, and those who could invoke it).

having been granted immunity from prosecution in what are known as 2511 letters.²³¹

Finally, amid growing calls to rethink or repeal Section 230, immunity for partnering with law enforcement may become more significant. While platforms' "right to exclude" user-generated content has increasingly been called into question, the obligation to comply with law enforcement demands remains at the core of many proposals to redesign Section 230.²³² If Congress were to enact legislation that renders platforms immune from suit when they take down "unlawful" content, they may also immunize platforms from liability for their collaboration with law enforcement in determining whether content is, in fact, unlawful.

V. CONCLUSION

Amid broadening recognition that social media platforms aggravate offline harms, like election tampering, communal violence, public health risks, and genocide, platforms' collaborations with law enforcement institutions bring both positive and negative effects. Synergies between public policy and private platform decision-making surely strengthen the government's ability to put its priorities into action. At the same time, however, the emergent ecosystem of information-sharing, collaboration, and public-private cooperation undermines the conventional wisdom that platform power necessarily comes at the expense of state authority, and vice versa. Law enforcement exerts both direct and indirect pressure on platform content rules, urging platforms to adopt more restrictive community standards, facilitate speedier takedowns, and share more information about harmful content with regulators. At the same time, platforms' affordances are reshaping law enforcement investigations and advancing surveillance.

Instead, as Reidenberg recognized, *lex informatica* can advance the goals of regulation as easily as inhibit them. The truth is that platforms in many contexts reflect the values of governments and specifically reflect the need for effective law enforcement. The same features that make social media so

231. Janus Kopfstein, *AT&T Getting Secret Immunity from Wiretapping Laws for Government Surveillance*, THE VERGE (Apr. 24, 2013, 2:42 PM), <https://www.theverge.com/2013/4/24/4261410/att-getting-secret-wiretapping-immunity-government-surveillance>.

232. Biden v. Knight Inst., 593 U.S. ___, 8 (2021) (Thomas, J., concurring); see Stop the Censorship Act, H.R. 4027 (116th Cong.), Sec. 2 (eliminating platforms' immunity for moderating content that it deems objectionable but preserving immunity for taking down "unlawful content"); Protecting Constitutional Rights from Online Platform Censorship Act, H.R. 83 (117th Cong.), Sec. 2 (making it unlawful for platforms to moderate "protected" content, and by implication excluding illicit material from the definition of "protected").

compelling as a technology of mass communication—the ability to instantaneously reach a broad audience—make it equally compelling as a technology of surveillance that is easy and cheap to use in both targeted and dragnet investigative contexts.

Yet twenty-three years after Reidenberg's germinal observations, U.S. law has made little progress in ensuring that *lex informatica* is as democratically legitimate or accountable as its regulatory equivalents. As technology firms increasingly rely on automation and predictive technology to define the boundaries between lawful and unlawful speech, the private policies and techniques of platform governance are increasingly transmuted into public law enforcement institutions. Despite the blurry boundaries between firm and state, the laws of surveillance and information-sharing continue to recognize a sharp divide between public and private actors. Amid growing calls to fundamentally rethink, reshape, or abolish U.S. policing, we should reconsider how the law enables private sector firms to act as a force multiplier for law enforcement.