BERKELEY TECHNOLOGY LAW JOURNAL

VOLUME 40, ISSUE 4

2025

Pages

1077–1364

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

Berkeley Technology Law Journal
University of California
School of Law
3 Law Building

Berkeley, California 94720-7200
editor@btlj.org
https://www.btlj.org

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 online four times each year by the Regents of the University of California, Berkeley.

Form. The text and citations in the *Journal* conform generally to THE CHICAGO MANUAL OF STYLE (18th ed. 2024) and to THE BLUEBOOK: A UNIFORM SYSTEM OF CITATION (Columbia Law Review Ass'n et al. eds., 22nd ed. 2025). Please cite this issue of the *Berkeley Technology Law Journal* as 40 BERKELEY TECH. L.J. ___ (2025).

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 the *BTLJ Podcast*, a series of episodes discussing new developments in law and technology.

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 Submissions may include previously unpublished articles, essays, book reviews, case notes, or comments concerning any aspect of the relationship between technology and the law.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., 22nd ed. 2025).

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.

SPONSORS
2024–25

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

BAKER BOTTS L.L.P.

COOLEY LLP

DAVIS POLK & WARDWELL LLP

DEBEVOISE & PLIMPTON LLP

DESMARAIS LLP

FENWICK & WEST LLP

FISH & RICHARDSON P.C.

GIBSON, DUNN & CRUTCHER LLP

GOODWIN PROCTER LLP

GUNDERSON DETTMER STOUGH VILLENEUVE FRANKLIN & HACHIGAN LLP

HAYNES AND BOONE, LLP

IRELL & MANELLA LLP

JONES DAY

KEKER, VAN NEST & PETERS LLP

KILPATRICK TOWNSEND & STOCKTON LLP

KIRKLAND & ELLIS LLP

KNOBBE, MARTENS, OLSON & BEAR, LLP

LATHAM & WATKINS LLP

MCDERMOTT WILL & SCHULTE LLP

MORGAN LEWIS & BOCKIUS LLP

MORRISON & FOERSTER LLP

ORRICK HERRINGTON & SUTCLIFFE LLP

PAUL HASTINGS LLP

QUINN EMANUEL URQUHART & SULLIVAN, LLP

ROBINS KAPLAN LLP

ROPES & GRAY LLP

SIDLEY AUSTIN LLP

SKADDEN, ARPS, SLATE, MEAGHER & FLOM LLP

WEIL, GOTSHAL & MANGES LLP

WHITE & CASE LLP

WILLKIE FARR & GALLAGHER LLP

WILSON SONSINI GOODRICH & ROSATI

WILMER CUTLER PICKERING HALE AND DORR LLP

WOMBLE BOND DICKINSON LLP

BOARD OF EDITORS
2024–25

Executive Board

BOARD OF EDITORS
2024-25

Editorial Board

Articles Editors

ALYSON CHIE
REBEKAH ENT
HAO HUANG
MONICA JEUNG
SRISHTI KHEMKA
JOSHUA LEE
MARIA MILEKHINA
LEA MOUSTAKAS
AARON MOSS
EMILY REHMET
JACOB SHOFET
DANIEL WARNER

Submissions Editors
JOSEPH KYBURZ
LILLY MAXFIELD
SARAH SISNEY
AMANDA TODD

Production Editors
SEUNGHAN BAE
MEGAN BIRMINGHAM
GLORIA CHAO
MICHELLE CHENG
SIENNA HORVATH
ALEX LE
TUONG-VI NGUYEN

Technical Editors
RYAN HAYDEN
YASAMEEN JOULAEE
STEPHANIE MORGAN
BEN PEARCE
FAYE ZOU

Notes & Comments Editors
ANDRA CERNAVSKIS
ERIC RITTER

Member Relations Editors
ANTONY NOVAK
TERRY ZHAO

Symposium Editors
NADIA GHAFFARI
NIYATI NARANG

Senior Podcast Editors
JULIETTE DRAPER
MEG O'NEILL

Podcast Editors
BRAXDON CANNON
JOY FU
LUCY HUANG
PAUL WOOD

External Relations Editors
MICHELLE D'SOUZA
LAUREL MCGRANE
JESSE WANG

Student Publications Editor
MAYA DARROW

LLM Editors
AIMILIA KECHAGIA
CARLOS NEYRA

Web & Technology Editors
KARINA SANCHEZ
EMILY WELSCH

MEMBERSHIP
2024-25

Members

RAQUEL ACOSTA	NOOSHA ALIABADI	BRYAN AHN
DEEMA AFANA	EMMA ALLEN	AYESHA ASAD
AVYSSA ABOUTORABI	RYAN AZIMI	PB BATLER
DAVID BERNSTEIN	ISABELLE BORCHARDT	QUINN BRASHARES
ZACH BROUSSEAU	JOSH BUCK	JOSE CAMACHO
DESHAWN CARTER	JOSE MIGUEL CATEPILLAN	JOSHUA CENZANO
MATT CHENG	JEFF CHEN	ALEXA CHAVARA
PENSIT CHAROANVIRIYAPAP	ZAHRA CHAUDHARY	SHRUTHI CHANDRAN
THANASIS CHRISTOU	SKYLAR CUSHING	BRADY COULTER
JAMES COURSER	RACHEL DAMES	FRANCO DELLAFIORI
MANUSHRI DESAI	NITZA DIAZ	VIKTOR DIMAS
KOSHA CHETAN DOSHI	JINYANG DU	ARTHUR FILPPU
JONAS EICKENBERG	MAX SYLVESTER EISELE	ELENA FABIAN
BECCA FRISCHLING	KAIMING GAO	MATTHEW GAVIETA
NEETI GANJUR	JENNIFER GHOSHRAY	PHILIP HUECK
JEFFREY GREER IV	PETER GUTHRIE	VASANTHI HARIHARAN
WHITNEY HARRIS	DAKSHINA HAZARIKA	SCARLETT HU
NATALIE HELLER	TAJA HIRATA-EPSTEIN	CHRISTOPHER HONG
DANIEL HONG	JACQUELINE HSING	ANAGHA IYER
ALLEN JI	VARSHA JHAVAR	CORINNE JOHNSTON
ANANYA KANORIA	SAMEER KAZIM	BRADLEY KHANTHAPHIXAY
TEYA KHALIL	JEYHUN KHALILOV	MADDISON KONWAY
LILY KRAFT	APARNA KUMAR	BRIAN LAI
GAURAV LALSINGHANI	MITCHELL LEE	KUNYU LI
DANPING LI	KARISSA LIN	SONGLIN LIU

MEMBERSHIP
2024-25

Members (continued)

APRIL LIU	HANNAH LONDON	JULIO LORET DE MORA
TOMASZ MACIAK	YANG MA	WINNIE MA
KRYSTLE MALIXI	SAUNTHARYA MANIKANDAN	VIRKEIN MANGLIK
SHREYA MAZUMDAR	JOSHUA MCDANIEL	ZAC MCPHERSON
ERIC MIRANDA	SASMITA MISHRA	ANNIE MOSKOFIAN
GRACE MURPHY	RAMESES NEALE	NICHOLAS NAVARRO
WEI NG	ILAMOSI OGWEMOH	VICTORIA ORINDAS
BRAYAN ORTIZ RAMOS	GAURAV PANDE	ALLEN PARK
GAVIN PATCHET	NANDITHA PILLAI	JENNY POWER
ZACHARY PRICE	MICHELLE PYKE	ARMAN RAMEZANI
BARBARA RASIN	VISHAL RAVI	LIAM ROCHE
BEATRIZ SAMPAIO	MAYA SAIDI LITVAK	ROCHELLE SIMS
JAMILE SIMAO	XIWEN SHEN	DEIRDRE SHEPARD
ARSENY SHEVELEV	GEORGY SHEVELEV	RUOCHENG SHI
ISHIKA SINGHAL	HAILEY STEWART	HANANYA SUNDERRAJ
RYAN SWARDSTROM	ANDREIA TAMASHIRO	CHLOE TEPPER
MALIA THORNTON	ZAK TURNER	KHUE TRAN
SHRIYA VERMA	MADHUNIKA VARADARAJAN	JEFF WANG
PINYAN WANG	RACHEL WANG	ELLEN WALLER
SHASWAT WEERAMONGKOLKUL	SIDNEY WRIGHT	ELLIOTT WILLIAMS
AVA WU	LINJING XIE	XUAN XIE
SIMON XU	CHIEH-HSIN YANG	ELOISE YANG
JASMINE YI	WILLIAM YAO	LIWEI YE
KAIDI ZHANG	NINA ZHANG	YUCHEN ZHAI
MICOL ZHAI	ALEXANDRA ZALEWSKI	KARI ZIMMERMAN

BERKELEY CENTER FOR LAW & TECHNOLOGY
2024-25

WAYNE STACY
Executive Director

Staff

ALLISON SCHMITT
Director, BCLT Life Sciences Law & Policy Center and Senior Fellow

JANN DUDLEY
Associate Director

RICHARD FISK
Assistant Director, Events & Communications

JUSTIN TRI DO
Media Coordinator

ABRIL DELGADO
Events Specialist

ALEXIS GOETT
Office Administrator

Fellow

RAMYA CHANDRASEKHAR
Biometric Regulatory Fellow

YUAN HAO
Senior Fellow

KATHRYN HASHIMOTO
Copyright Law Fellow

MIRIAM KIM
Practitioner Fellow, Generative AI

ROBERT BARR
BCLT Executive Director Emeritus

BERKELEY CENTER FOR LAW & TECHNOLOGY
2024-25

Faculty Directors

KENNETH A. BAMBERGER
Rosalinde and Arthur Gilbert Foundation Professor of Law

ELENA CHACHKO
Assistant Professor of Law

COLLEEN CHIEN
Professor of Law

CATHERINE CRUMP
Director, Samuelson Law, Technology & Public Policy Clinic and Clinical Professor of Law

CATHERINE FISK
Barbara Nachtrieb Armstrong Professor of Law

CHRIS JAY HOOFNAGLE
Professor of Law in Residence

SONIA KATYAL
Roger J. Traynor Distinguished Professor of Law

ORIN S. KERR
William G. Simon Professor of Law

PETER S. MENELL
Koret Professor of Law

ROBERT P. MERGES
Wilson Sonsini Goodrich & Rosati Professor of Law

DEIRDRE K. MULLIGAN
Professor in the School of Information and School of Law

TEJAS N. NARECHANIA
Robert and Nanci Corson Assistant Professor of Law

BRANDIE NONNECKE
Associate Research Professor at the Goldman School of Public Policy

OSAGIE K. OBASOGIE
Haas Distinguished Chair and Professor of Law

ANDREA ROTH
Professor of Law

PAMELA SAMUELSON
Richard M. Sherman Distinguished Professor of Law and Information

PAUL SCHWARTZ
Jefferson E. Peyser Professor of Law

ERIK STALLMAN
Associate Director, Samuelson Law, Technology & Public Policy Clinic and Assistant Clinical Professor

JENNIFER M. URBAN
Director, Samuelson Law, Technology & Public Policy Clinic and Clinical Professor of Law

MOLLY SHAFFER VAN HOUWELING
Harold C. Hohbach Distinguished Professor of Patent and IP Law

REBECCA WEXLER
Assistant Professor of Law

FOREWORD: AI GOVERNANCE AT THE CROSSROADS

Emily Rehmet[†] *& Yasameen Joulaee*[††]

Artificial intelligence (AI) is no longer a promise of the future. Rather, it is a technological force actively reshaping civil society and private institutions, "sinking deeper and deeper into the infrastructures of our everyday life."[1] The rapid evolution of AI signals the dawn of another technological revolution—one that promises to be as impactful as the rise of the internet, if not more.

But this newfound technological power brings with it a sense of profound responsibility. As AI systems become increasingly sophisticated, there is an urgency to ensure safe and responsible public and private use. Policymakers, regulators, and legal scholars face a pivotal question at a critical juncture in technological advancement: how should we guide the deployment of AI in ways that simultaneously maximize innovation and protect individual fundamental human rights, democratic institutions, and the general public welfare? The Berkeley Center for Law and Technology and the *Berkeley Technology Law Journal*'s 28th Annual Spring Symposium, "AI Governance at the Crossroads," aimed to answer this question.

On February 27 and 28, 2025, scholars, policymakers and leading AI-industry participants from the United States and abroad convened at the University of California, Berkeley, School of Law to discuss initiatives to

DOI: https://doi.org/10.15779/Z38PR7MX1S

† Co-Editor-in-Chief 2025–26, Berkeley Technology Law Journal; J.D. Candidate, University of California, Berkeley, School of Law, 2026.

†† Co-Editor-in-Chief 2025–26, Berkeley Technology Law Journal; J.D. Candidate, University of California, Berkeley, School of Law, 2026. Our heartfelt appreciation to Professor Molly Van Houweling for her kind feedback and encouragement.

1. Professor Mona Sloane advanced a sociological approach to AI governance, describing contemporary AI systems as "sinking deeper and deeper into the infrastructures of our everyday life" and, in many respects, "already . . . social infrastructure because it's everywhere we look." Mona Sloane, Panelist, Federal Approaches to AI Governance Panel at the Berkeley Technology Law Journal's 28th Annual Spring Symposium: AI Governance at the Crossroads (Feb. 28, 2025). In her corresponding Article, Professor Sloane examines how digital technologies have become embedded in social infrastructure through a case study of AI use in classrooms. Mona Sloane, Ella Duus & Bertrall Ross, *Students, Power, and Technology*, 40 BERKELEY TECH. L.J. 1315 (2025).

regulate artificial intelligence at the local, state, national,[2] and international[3] levels. The first day of the conference, the "Tutorial on AI Governance Basics" provided a technical tutorial on AI and introduced major state, federal, and international AI governance initiatives. The second day of the symposium explored perspectives on a range of AI policies and regulations. Scholars and practitioners offered assessments of how well a range of policy proposals would promote artificial intelligence innovation and development, benefit the public good through such technologies, and ensure safe and trustworthy systems.

The 40.4 Symposium Issue of the *Berkeley Technology Law Journal* herein synthesizes these proposals and provides a written forum authored by the leading voices from academia and government to further explore this moment in technological innovation. The articles comprising this Issue critically examine the legal, institutional, and ethical frameworks needed to appropriately govern frontier AI technologies. The authors propose ways in which governance tools can be properly wielded to guide innovation, mitigate AI harm,[4] and provide for accountability when such harm occurs.

Implementing governance that ensures responsible and safe AI use, while also fostering entrepreneurship and efficiency, is an exceptionally complex undertaking—one that demands sustained, thoughtful leadership from the public sector,[5] active engagement from civil society,[6] and robust support from the private sphere.[78] The diverse perspectives shared by industry leaders, domestic policymakers, technologists, social scientists, and legal scholars

2. Sorelle A. Friedler & Andrew D. Selbst, *The OMB Artificial Intelligence Memoranda*, 40 BERKELEY TECH. L.J. 1237 (2025).

3. Margot E. Kaminski & Andrew D. Selbst, *An American's Guide to the EU AI Act*, 40 BERKELEY TECH. L.J. 1081 (2025); Isabela Ferrari, Niyati Narang & Colleen Chien, *JudgeGPT: When Progress Meets Precedent*, 440 BERKELEY TECH. L.J. 1185 (2025).

4. Friedler & Selbst, *supra* note 2.

5. Deirdre K. Mulligan & Kenneth A. Bamberger, *Recentering Public Values in AI Governance: Examples from the Biden Administration*, 40 BERKELEY TECH. L.J. 1135 (2025).

6. Professor Sloane warns that a concurrent growth of digital technology on campus and a decline in student civil engagement in university governance is troubling. In turn, she urges the establishment of student technology councils that advise faculty and administration to position students as co-governors. Sloane et al., *supra* note 1.

7. Alexander R. Mueller & Christopher S. Yoo, *Taking Standards Seriously: The Case for a Private Standards-Based Approach to AI Governance*, 40 BERKELEY TECH. L.J. 1273 (2025).

8. As the panel's final speaker, Alla Seiffert of Amazon Web Services offered an industry perspective on AI governance. Drawing on her expertise in government procurement and her senior role on Amazon Web Services's Public Policy team, she addressed the effects of government AI directives on private vendors such as Amazon. Alla Seiffert, Panelist, Federal Approaches to AI Governance Panel at the Berkeley Technology Law Journal's 28th Annual Spring Symposium: AI Governance at the Crossroads (Feb. 28, 2025).

during the "AI Governance at the Crossroads" Symposium converge on a clear takeaway: no single sector can, or should, shoulder this responsibility alone. With the shift to a deregulatory regime under the Trump Administration's AI Action Plan, it will be ever-more important to consider the Symposium's opening words that the private sector, nonprofit institutions, civil society, and the government will all need to play a part.[9] The six articles collected in this Issue lay the groundwork to unite for that collaboration.

We are deeply grateful to Symposium Editors Niyati Narang and Nadia Ghaffari, whose vision and leadership made this symposium possible; to our distinguished symposium speakers, whose insights shape national and international dialogue on such transformative technology; and to the authors in this volume, whose care and compassion reflect a shared commitment to safeguarding the future safety and livelihood of the public.

We hope this issue serves not only as a reflection of where we stand today, but as a call to action. The policy choices we make now will shape the trajectory of AI for decades, if not the next century, to come. Thoughtful and forward-looking AI governance is not just a political challenge—it is a democratic imperative.

9. In her opening keynote address, Professor Deirdre Mulligan urged that "not just the private sector, but nonprofits, civil society, and government all need to have a hand" in AI governance. Deirdre Mulligan, Panelist, Opening Keynote at Berkeley Technology Law Journal's 28th Annual Spring Symposium: AI Governance at the Crossroads (Feb. 28, 2025); Mulligan & Bamberger, *supra* note 5.

AN AMERICAN'S GUIDE TO THE EU AI ACT

Margot E. Kaminski† *& Andrew D. Selbst*††

The EU AI Act entered into force in August 2024. The AI Act is long. It is complicated. It relies on a regulatory framework and institutions unfamiliar to many in the United States. But as the first omnibus AI regulation worldwide, it has the potential to have a vast influence on both practice and lawmaking.

In this Article, we provide the American's Guide to the EU AI Act. This Article breaks down the AI Act for a U.S. law audience, explaining the overall mechanisms, and how the Act interacts with background EU laws and institutions. At its core, the AI Act is structured on Europe's product safety regime. It is aimed at governing AI systems through assigning them into risk tiers, and deploying bans, risk regulation, and self-regulation. But it also contains later-drafted provisions on general-purpose AI that depart from this framework, as well as multiple ad-hoc provisions and other regulatory strands.

The Article also analyzes the consequences of framing AI regulation as risk regulation and of constructing AI systems as products rather than bureaucratic processes. It describes the AI Act itself as a "legal exoskeleton," with hard law built around the softer belly of technical standards. The Article identifies the threats this poses for both substance and legitimacy, and the potential political ramifications of that design.

TABLE OF CONTENTS

DOI: https://doi.org/10.15779/Z38JS9HB12

† Moses Lasky Professor of Law, Colorado Law School, and Director of the Privacy Initiative at Silicon Flatirons Center. Recipient of a 2024 Fulbright-Schuman grant and Fernand Braudel Senior Fellowship at the European University Institute (EUI). Thanks to Deirdre Curtin for hosting and involving Professor Kaminski in the EUI intellectual community and related discussions. Thanks to Marco Almada and Nicolas Petit for detailed and thoughtful feedback on this piece. Mistakes are all ours.

†† Professor of Law, University of California, Los Angeles, School of Law.

I. INTRODUCTION

The EU AI Act entered into force in August 2024. The AI Act is long. It is complicated. It relies on a regulatory framework and institutions unfamiliar to many in the United States. But as the first omnibus AI regulation worldwide, it has the potential to have a vast influence on both practice and lawmaking.

In this Article, we provide an American's Guide to the EU AI Act. We begin in Part II with necessary background on EU law and institutions that readers can skim or skip, if they are already familiar with the European Union (EU). In Part III, we give an overview of the AI Act. At its core, the AI Act is structured on Europe's product safety regime. It is aimed at governing predictive AI systems through assigning them into risk tiers, and deploying bans, risk regulation, and self-regulation. But it also contains later-drafted provisions on general-purpose AI that depart from this framework, as well as multiple ad hoc provisions.

In Part IV, we offer our analysis. We point to the consequences of framing AI regulation as risk regulation and of constructing AI systems as products rather than bureaucratic processes. We describe the AI Act itself as a "legal exoskeleton," with hard law built around the softer belly of technical standards. We identify the threats this poses for both substance and legitimacy, and the potential political ramifications of that design.

We close in Part V with a word of warning about global power politics. This may not be the time for another legal export out of Brussels. And the AI Act itself is not well-designed to be exported. Instead, we caution that deregulatory forces are compounding, both within Europe and from the United States.

II. SOME NECESSARY BACKGROUND ON EU INSTITUTIONS, EU LAW, AND MEMBER STATE INSTITUTIONS

The AI Act sits atop a mountain of existing EU law and institutional structure. In this section, we begin with some basic and not-so-basic background that we think is necessary to understand what's going on with the AI Act.

A. EU INSTITUTIONS

We start with EU institutions, briefly covering both the basics and how these institutions are relevant to the AI Act. The EU is a supranational organization consisting of a system of twenty-seven member states, joined in a common legal and economic enterprise. An American audience might understand it as a kind of federalist system, though perhaps one more akin to the one imagined by the Articles of Confederation, with member states exercising true sovereignty.

Certain lawmaking institutions operate at the EU level by either directly legislating or delegating to member states. The EU has four primary decision-making bodies: the European Council, the European Parliament, the Council of the European Union, and the European Commission. The European Council—not to be confused with the Council of the EU[1]—comprises the EU heads of state. It sets the general political direction of the EU but is not generally involved in legislation. The two main legislative bodies of the EU are the European Parliament and the Council of the European Union. The Parliament represents the citizens of member states and is directly elected by them. The Council of the EU consists of ministers from each member state and differs depending on the policy area of the law being considered. The primary executive body of the EU is the European Commission.

We offer the following overview of the EU legislative process because the AI Act in several places envisions bypassing it for purposes of amendments or

1. It's also important not to confuse either with the "Council of Europe," a totally separate international institution comprising 46 countries, including all EU members, and dedicated to upholding human rights and democracy. We know this is a big ask, because wow, that's a lot of different things named the "Council." But, you know, just try.

implementation. As a general matter, new EU laws are typically passed through what is known as the "ordinary legislative process," in which the Commission uses its "right of initiative" to propose a new law, which is then taken up jointly by the Parliament and Council of the EU. A new law will often undergo a lengthy "trilogue" negotiation, an informal institutional negotiation in which members of the three bodies come together to work out a draft, which can then be formally adopted by each of the three bodies internally. This is not unlike a bill going to conference in the U.S. Congress, except more complicated, as there are three bodies that need to agree. This is the process that the AI Act followed in its three-year drafting history.

The Commission (executive) has certain specific roles designated by the AI Act, including the delegated ability to modify certain aspects of the law without going through the full legislative process.[2] The Act also created an "AI Office" within the Commission to "develop Union expertise and capabilities in the field of AI."[3] This European AI Office has already been involved, primarily as a convener, in drawing up a General-Purpose AI Code of Practice.[4] The AI Office has also released a template for summarizing training data for general-purpose AI.[5]

Back to EU law: the powers of the EU are conferred by treaty. Unlike EU member states whose sovereignty is assumed, the European Union does not have inherent powers, because it is not a sovereign state.[6] The EU's power to

2. Regulation (EU) 2024/1689 of the European Parliament and of the Council, art. 97, 2024 O.J. (L 2024/1689) [hereinafter AI Act].

3. *Id.* art. 64.

4. *Drawing-Up a General-Purpose AI Code of Practice*, EUR. COMM'N (Aug. 1, 2025), https://digital-strategy.ec.europa.eu/en/policies/ai-code-practice ("The AI Office played a pivotal role throughout the process . . . facilitating the drawing-up, coordinating the discussions and documenting the outcomes"). On August 1, 2025, the "Commission and AI Board approve[d] the code via Adequacy Decisions."

5. *See* EUROPEAN AI OFFICE, EUROPEAN AI OFFICE WORKING GROUP MEETINGS, CODE OF PRACTICE FOR GENERAL PURPOSE AI: TEMPLATE FOR SUMMARY OF TRAINING DATA (Jan. 17, 2025), https://ec.europa.eu/newsroom/dae/redirection/document/111909; *see also Drawing-Up AI Code of Practice*, *supra* note 4 ("[T]he AI Office is also developing a template on the sufficiently detailed summary of training data that general-purpose AI model providers are required to make public according to Article 53(1)d) of the AI Act . . . . [T]he AI Office has presented its preliminary ideas and allowed the participants to the Code to provide additional feedback on the preliminary structure and elements of the template . . . [and] was also discussed with the Member States' representatives in the AI Board subgroup and the European Parliament before the Commission adopts the template in the second quarter of 2025.").

6. This principle is referred to as the "principle of conferral." *See* Consolidated Version of the Treaty on European Union art. 5(2), June 7, 2016, 2016 O.J. (C 202) 13 [hereinafter TFEU] ("Under the principle of conferral, the Union shall act only within the limits of the competences conferred upon it by the Member States in the Treaties to attain the objectives

make law—referred to as its "competences"—varies in different subject matter areas, depending on what treaties confer to it.[7] In some areas, the EU's power is exclusive, while in others it is shared with or supporting of the member states.[8] For example, the EU has exclusive power over the Euro and governing trade; shared power over consumer protection, some types of social policy, and governance of the internal market more generally; and supporting power only for areas such as industry, culture, and tourism.[9] Member states retain power for internal security and crime prevention,[10] and in general, any power not explicitly conferred on the EU belongs to member states.[11]

While this "principle of conferral" formally still governs, it does seem that the EU is undergoing an expansion of authority similar to the U.S. federal government's under the Commerce Clause.[12] Many different legal issues are considered regulation of the "internal market," and thus can be subject to EU power.[13] The AI Act, as explained below, is part of a larger EU product safety regime justified as a regulation of the internal market.[14]

---

set out therein. Competences not conferred upon the Union in the Treaties remain with the Member States.").

7. *See id.* title I, art. 2.

8. *Id.* art. 2(1) ("When the Treaties confer on the Union exclusive competence in a specific area, only the Union may legislate and adopt legally binding acts, the Member States being able to do so themselves only if so empowered by the Union or for the implementation of Union acts."); *id.* art. 2(2) ("When the Treaties confer on the Union a competence shared with the Member States in a specific area, the Union and the Member States may legislate and adopt legally binding acts in that area."); *id.* art. 2(5) ("In certain areas and under the conditions laid down in the Treaties, the Union shall have competence to carry out actions to support, coordinate or supplement the actions of the Member States, without thereby superseding their competence in these areas.").

9. *See generally infra* Part III; *see also* TFEU, *supra* note 6.

10. *See* TFEU, *supra* note 6, art. 72 ("This Title shall not affect the exercise of the responsibilities incumbent upon Member States with regard to the maintenance of law and order and the safeguarding of internal security."); *see also* TFEU, *supra* note 6, arts. 275–76 (describing the limits of the jurisdiction of the European Court of Justice).

11. *See id.* arts. 4(1), 5(2); Sacha Garben, *Competence Creep Revisited*, 57 J. COMMON MKT. STUD. 205, 205 (Mar. 2019) ("[P]owers that have not been explicitly conferred on the EU remain exclusively with the Member States (Article 4(1) and 5(2) TFEU).").

12. *See, e.g.*, ROBERT SCHUTZE, INTRODUCTION TO EUROPEAN LAW 64 (4th ed. 2023) ("[T]hree developments have led to widespread accusations that the European Union's competences are 'unlimited.'").

13. Garben, *supra* note 11, at 207 ("The Treaty's functional powers—mostly, but not exclusively related to the internal market—can cut horizontally through all policy areas, including those where the EU has no, or only complementary, competence. This means that the EU can, through such indirect powers, legislate in areas that are considered to fall within national autonomy.").

14. *See* AI Act, *supra* note 2, art. 1 (describing the purpose of the regulation as, among other things, "to improve the function of the internal market.").

The principal EU court is the Court of Justice of the European Union (CJEU).[15] The CJEU works in parallel with national courts of member states to effectuate EU law.[16] One of its main functions is to interpret EU law. The CJEU is one institution but has two court with different jurisdiction: the European Court of Justice (ECJ) and the General Court. The ECJ primarily hears cases referred to them by the high courts of member states, including preliminary rulings on new issues of law. The General Court, by contrast, can hear directly from individuals and institutions, but only for claims that an EU body directly violated their rights.

Generally, if an individual wants to raise a claim that private parties or national institutions have violated EU law, she must go through her national court.[17] National courts hear disputes on the application of EU law to their citizens and can refer an issue of European law to the ECJ. This is referred to as the "preliminary reference procedure."[18] The role of the ECJ can vary depending on the structure of a particular treaty. The ECJ has an active role, for example, in data protection law. Under the General Data Protection Regulation,[19] a data subject can file a complaint with the local data protection authority, and if dissatisfied, can sue that authority in national court with the ECJ as a potential backstop. But even under the GDPR, which grants data subjects many individual rights, a data subject may not sue in the ECJ directly.

By contrast, although the AI Act is purportedly motivated in large part by concern for individual fundamental rights, there is no individual right of

15. Not to be confused with the European Court of Human Rights, the court that oversees the European Convention on Human Rights, a treaty passed by the Council of Europe. *See generally* ROBERT SCHUTZE, EUROPEAN UNION LAW, ch. 10: Judicial Powers I: (Centralized) European Procedures (3d ed. 2021).

16. SCHUTZE, *supra* note 15, ch. 10 ("In addition to a number of direct actions (direct actions start directly in the European Court), the EU Treaties here envisage an indirect action starting in the national courts: the preliminary reference procedure. This procedure is the judicial cornerstone of the Union's cooperative federalism. For it combines the central interpretation of Union law by the Court of Justice with the decentralised application of European law by the national courts.").

17. Individuals may bring an "action for annulment" of EU law under art. 263 of TFEU, asking for a particular EU law to be annulled. *See* Action for Annulment, EU: Summaries of EU Legislation (Oct. 29, 2010), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=legissum:ai0038. However, the Court favors the preliminary reference procedure. *See* SCHUTZE, *supra* note 15, at 373. And if an individual wants to challenge a purported violation of EU law, they use preliminary reference.

18. *See* TFEU, *supra* note 6, art. 267(1)(b); *see also* Preliminary Ruling Proceedings–Recommendations to National Courts, EU: Summaries of EU Legislation (Dec. 3, 2024), https://eur-lex.europa.eu/EN/legal-content/summary/preliminary-ruling-proceedings-recommendations-to-national-courts.html.

19. Regulation (EU) 2016/679 of the European Parliament and of the Council, 2016 O.J. (L 119) 1 [hereinafter GDPR].

redress in the Act. The institutions involved are focused on the operation of products in the market rather than hearing cases on fundamental rights. Thus, the role of the CJEU in performing oversight or backstopping much of the law is less clear.[20]

We have thus far covered the main lawmaking, executive, and judicial institutions of the EU. The AI Act also refers to several other institutions that may be unfamiliar to a U.S. audience. The AI Act relies on the European Data Protection Supervisor (EDPS), a body that exists to enforce data protection law, to serve a few different functions. The European Data Protection Supervisor is designated as the overseer where an EU-level body acts in a capacity that would normally be overseen by a member state institution such as a market surveillance authority.[21] It is also designated as an observer for the EU AI Board created by the law.[22]

Then there are the European Standards Organizations (ESOs). These are private organizations, not governmental, but they are considered European bodies and are governed by EU law. There are three standards bodies authorized by certain EU laws to create technical standards that help effectuate the AI Act. Each has a slightly different area of expertise. While the European Committee for Standardization (CEN) is more generalized, the European Committee for Electrotechnical Standardization (CENELEC) focuses on standards related to electrical engineering, and the European Telecommunications Standards Institute (ETSI) focuses on information and communications.[23] The AI Act calls for and heavily relies on "technical" standards development, and has enlisted a joint technical committee of CEN and CENELEC to implement the AI Act's standards.[24]

B. EU LAWS

Here, we turn from EU institutions to EU laws. We first cover the meaning of commonly used terms, like regulations, directives, and recitals. Then we turn to the broader substantive legal setting behind the AI Act, identifying several relevant and overlapping EU laws.

EU statutory laws come in two principal forms: regulations and directives. A regulation is a law that is directly binding on individual entities of the EU—

20. Some of the provisions of the AI Act may be found to have direct effect and thus be subject to being raised before national courts. Thanks to Nicholas Petit for this important caveat.

21. AI Act, *supra* note 2, art. 74(9).

22. *Id.* art. 65.

23. *European Standardization*, CEN-CENELEC, https://www.cencenelec.eu/european-standardization/cen-and-cenelec/.

24. *Artificial Intelligence*, CEN-CENELEC, https://www.cencenelec.eu/areas-of-work/cen-cenelec-topics/artificial-intelligence/.

citizens, companies, and member states. A directive is, by contrast, a law that directs member states to pass laws on a topic, within certain parameters. If you think of the EU as supranational, then regulations are akin to self-executing treaties, while directives are akin to non-self-executing treaties. A principal function of EU law in general is the harmonization of member state laws. A directive is used when the EU feels that a lighter touch on harmonization is required, whereas a regulation is passed to achieve tighter harmonization.

For example, the GDPR was passed in 2016 because the data protection laws of member states under the 1995 Data Protection Directive (DPD)[25] were widely fragmented as implemented.[26] Also, a directive will sometimes serve as a stepping stone to a later regulation if deemed necessary by the legislative bodies, as in the transition from the DPD to the GDPR, or in the case of the e-Commerce Directive[27] and the Digital Services Act.[28]

EU laws are made up of articles, recitals, and sometimes annexes. The articles make up the binding legislative text. Recitals, by contrast, illustrate the purpose behind the law; they are like formalized legislative history. But unlike legislative history in the United States—which is often denigrated as either a last resort or irrelevant to interpretation—recitals are vitally important. The CJEU operates under a theory of purposive interpretation, in which text is to be interpreted in light of its declared purpose.[29] Thus, recitals play a direct role in giving meaning to EU law.

Some EU laws contain annexes, which are also binding legislative text, written separately. An article may refer to an annex for a procedure or list of covered circumstances to apply a particular provision. The separation allows for easier updates of implementing details. Annexes are heavily used in the AI Act.

Now we turn from structure to legal substance. The AI Act implicitly and explicitly relies on, overlaps with, and is constrained by other EU laws. The most relevant of these is a model known as the "New Legislative Framework" (NLF)—the framework for the EU's product safety regime. Laws built on the

25. Directive 95/46/EC, of the European Parliament and of the Council, 1995 O.J. (L 281) 31.

26. GDPR, *supra* note 19, recital 9 ("[O]bjectives and principles of Directive 95/46/EC remain sound, but it has not prevented fragmentation in the implementation of data protection across the Union.").

27. Directive 2000/31/EC, of the European Parliament and of the Council (Directive on electronic commerce), 2000 O.J. (L 178) 1.

28. Regulation (EU) 2022/2065, of the European Parliament and of the Council (Digital Services Act), 2022 O.J. (L 277) 1.

29. *See, e.g.*, Gerard Conway, *The Quality of Decision-Making at the Court of Justice of the European Union*, *in* HOW TO MEASURE THE QUALITY OF JUDICIAL REASONING 225, 227 (Mátyás Bencze & Gar Yein Ng eds., 2018).

NLF operate by requiring products to meet certain "essential requirements" for safety before entering the market.[30] They are allowed to enter the market only after a "conformity assessment" is performed, showing that the product—as well as its risk mitigation framework, failure alert procedures, and documentation—meet the framework's essential requirements. Typical NLF laws regulate children's toys[31] or elevators[32]—standard product safety regimes that might be assigned to the Consumer Product Safety Commission in the United States. Familiarity with the NLF is principally important because the AI Act was created as a product safety law under the NLF and uses the same conformity assessment and market surveillance oversight structures that the rest of the NLF laws use.[33] We go into more specifics below when we discuss the substance of the law.

Other laws are also important to the AI Act for different reasons. The Charter of Fundamental Rights of the European Union lays out the EU's fundamental rights regime.[34] The AI Act relies on the Charter and interpreting cases implicitly in two ways. First, the Act indicates that AI poses risks to fundamental rights, but does not delineate which rights are affected or how they might be affected by AI. It therefore leans on the existing EU fundamental rights framework to fill this gap. The Act also does not offer individual rights of redress, instead deferring to other existing laws to enable individual redress for violations of fundamental rights.

The data protection regime of the EU, specifically the GDPR, is a particularly important complement to the AI Act. While AI is built on data, the AI Act is not a data protection law. For that, there is the GDPR. Data protection is a fundamental right protected by the Charter.[35] The GDPR is backstopped by the CJEU, which interprets the GDPR's regulatory

30. Stéphane du Boispéan, Markus Mueck & Christophe Gaie, *Introduction to the European New Legislative Framework*, *in* EUROPEAN DIGITAL REGULATIONS 1, 2–3 (Markus Mueck & Christophe Gaie eds., 2025).

31. Directive 2009/48/EC of the European Parliament and of the Council (on the safety of toys), 2009 O.J. (L 170) 1.

32. Directive 2014/33/EU of the European Parliament and of the Council, 2014 O.J. (L 96) 251.

33. The NLF has also recently been applied to "products with digital elements" through the Cyber Resilience Act (Regulation (EU) 2024/2847 of the European Parliament and of the Council, 2024 O.J. (L. Series)). According to Marco Almada, in communications with the Authors, "[i]t does not face many of the issues created by the AI Act, as its regulatory object is much more technical." (communication on file with authors). But the Cyber Resilience Act, too, raises concerns about using a risk-based approach to fundamental rights. *See* Pier Giorgio Chiara, *Understanding the Regulatory Approach of the Cyber Resilience Act: Protection of Fundamental Rights in Disguise?*, 16 EUR. J. RISK REGUL. 469 (2025).

34. Charter of Fundamental Rights of the European Union, 2000 O.J. (C 364) 1.

35. *Id.* arts. 7, 8.

requirements in light of fundamental rights. Thus, even where the CJEU may appear to be absent from the AI Act, it is actively present in a closely complementary and overlapping law.

The GDPR and AI Act each seek to mitigate or protect against different types of harms to different sets of people: harms to data subjects versus harms to those affected by AI. However, some aspects of the laws directly overlap, for example where personal data is used to make decisions about individuals. Some requirements like the right to explanation of an AI decision compete or overlap.[36]

The AI Act also creates exceptions to some GDPR rules. For example, it allows the processing of "special category" data under the GDPR (i.e., sensitive data like race) "when strictly necessary for the purpose of ensuring bias detection and correction," as long as certain safeguards are followed.[37] The Act also permits the processing of personal data for regulatory sandboxing—essentially AI pilot programs with safe harbors.[38] Finally, as mentioned above, the Act designates roles for the European Data Protection Supervisor—an office created by the GDPR—including making the EDPS the relevant oversight authority for EU entities that would be regulated by the law.[39]

The AI Act also interacts substantially with the Digital Services Act. The DSA covers the use of platforms, and those that use AI for content moderation will be subject both to the DSA and the AI Act. In that case, the DSA is the *lex generalis* to the AI Act's *lex specialis*, and the platform AI will be governed by the DSA, with the AI Act's requirements riding atop it. The same principle governs AI that happens to also be a product covered by an existing NLF regime, like a toy or a drone. The AI Act instructs providers to comply with the existing NLF framework, but to add on the AI Act's requirements.[40] Finally, the Act also overlaps with EU copyright law, especially the Copyright in the Digital Single Market Directive.[41]

---

36. *Compare* Case C-203/22, CK v Magistrat der Stadt Wien, ECLI:EU:C:2025:117 (Feb. 27, 2025), *with* AI Act, *supra* note 2, art. 86; *see also* Margot E. Kaminski & Gianclaudio Malgieri, *The Right to Explanation in the AI Act*, *in* THE EU ARTIFICIAL INTELLIGENCE ACT: A THEMATIC COMMENTARY (Gianclaudio Malgieri, Gloria González Fuster, Alessandro Mantelero & Gabriela Zanfir-Fortuna eds., forthcoming 2026) (describing the "hydraulic effect" between AI Act's art. 86 and other law including the GDPR's art. 22).

37. AI Act, *supra* note 2, art. 10(5).

38. *Id.* art. 59(1); *see* Part II.

39. AI Act, *supra* note 2, arts. 70, 74.

40. *See id.* art. 43(3); *see also* Part II.

41. *See* João Pedro Quintais, *Generative AI, Copyright and the AI Act*, 56 COMPUT. L. & SEC. REV. 106107 (2025).

C. MEMBER STATE INSTITUTIONS

The Act also relies on several institutions belonging not to the EU, but to member states. Member states are instructed to "establish or designate as national competent authorities at least one *notifying authority* and at least one *market surveillance authority*."[42] Recall that the NLF product safety regime on which the Act is built relies on a conformity assessment process that in effect certifies compliance with the law before a product can move on the EU market. Sometimes, conformity assessments are conducted by third parties known as conformity assessment bodies, which are usually private organizations. "Notifying authorities" are member-state-level institutions that can set up procedures to certify conformity assessment bodies as "notified bodies."[43] A conformity assessment body can become a notified body by applying to the notifying authority and satisfying its procedure to check for competence.[44]

The second type of institution, the market surveillance authority, is typically a state agency that oversees product safety generally. The AI Act permits member states to designate other authorities than their existing product safety authorities as market surveillance authorities for purposes of the Act. As discussed below, these member-state-level institutions are important to the accountability frameworks for the Act, and central to the NLF as a whole.

The AI Act also relies on member state institutions that relate to other areas of law. The Act specifically relies on authorities charged with the enforcement of fundamental rights. It grants "the power to request and access any documentation created or maintained under this Regulation" to "[n]ational public authorities or bodies which supervise or enforce the respect of obligations under Union law protecting fundamental rights."[45] This description is vague in order to account for the variation in such bodies among member states, and as instructed by the AI Act,[46] most member states have designated which specific bodies meet that criterion.[47] To the extent the law intersects with data protection, member state data protection authorities will also have a role in enforcement.

---

42. AI Act, *supra* note 2, art. 70 (emphases added).
43. *Id.* art. 28(1).
44. *Id.* art. 29.
45. *Id.* art. 77(1).
46. *Id.* art. 77(2).
47. Poklaszlo, *Responsible Authorities for the Enforcement of the AI Act on National Level*, GDPRBLOG (Oct. 2, 2024), https://gdpr.blog.hu/2024/10/02/responsible_authorities_for_the_enforcement_of_the_ai_act_on_national_level (collecting the information).

III. THE LAW

The AI Act centrally aims to promote the uptake of AI throughout Europe, by mitigating risks from the use of AI systems. The very first article of the Act states: "[t]he purpose of this Regulation is to improve the functioning of the internal market and promote the uptake of human-centric and trustworthy artificial intelligence (AI), while ensuring a high level of protection of health, safety, fundamental rights . . . ."[48]

For readers outside of the EU accustomed to the stereotype of top-down centralized and heavy-handed regulation, the Act's explicit prioritization of market functioning and of the uptake of AI may be surprising. But the EU was initially a trading bloc, with its early laws concerned with opening up its common market.[49] Thus, to read the Act solely as an instantiation of Europe's adoption of the precautionary principle is a mistake. The Act is in large part concerned with clearing the way for the circulation and adoption of AI through the many member states of EU, through a uniform regulatory framework.[50] Whether that framework works or not is a different discussion.

The AI Act's regulatory framework is centrally concerned with mitigating risks.[51] The Act defines "risk" as "the combination of the probability of an occurrence of harm and the severity of that harm."[52] It attempts to mitigate multiple kinds of risks through a largely unified approach, categorizing AI into different "risk tiers." However, the types of risks the Act attempts to mitigate are varied: risks to "health, safety, [and] fundamental rights."[53] While this may appear to be a short list, the reference to "fundamental rights" opens up the Act's coverage considerably, to include every fundamental right referenced in the Charter.

This attempt to use the same framework for diverse types of risks creates problems. A quantitative framework focused on predicting and measuring potential harms can work well for some types of harms (e.g., to physical safety)

48. AI Act, *supra* note 2, art. 1(1).

49. The precursor to the EU was the European Economic Community (EEC), which began as a pact to regulate coal and steel. *See, e.g.*, *European Union: History*, GLOBALEDGE, https://globaledge.msu.edu/trade-blocs/european-union/history.

50. *See* AI Act, *supra* note 2, recital 1 ("This Regulation ensures the free movement, cross-border, of AI-based goods and services, thus preventing Member States from imposing restrictions on the development, marketing and use of AI systems, unless explicitly authorised by this Regulation.").

51. *See generally* Margot E. Kaminski, *Regulating the Risks of AI,* 103 B.U. L. REV. 1347 (2023).

52. AI Act, *supra* note 2, art. 3(2).

53. *Id.* art. 1(1).

but not for others (e.g., to human rights).[54] And the type of risk mitigation that might be appropriate for preventing physical harms might look quite different from how one ideally would regulate, for example, harms to privacy or free speech. Moreover, the Act's central reliance on a risk-mitigation (rather than rights-protective) framework arguably assumes AI adoption and that in the process of AI adoption some residual harms to fundamental rights are acceptable.

At its core—its regulation of "high risk" AI systems—the AI Act builds on the framework of EU product safety law. However, multiple aspects of the Act were added ad hoc and thus don't fit squarely into the product safety framework. Some were added at later stages of drafting. For example, after ChatGPT was publicly released during the Act's drafting process,[55] lawmakers came up with a distinct framework for regulating what the Act terms "general-purpose AI." Other ad hoc elements arose in response to feedback from various constituencies, including data protection regulators.[56]

Framing the substance of the Act in this way—(A) product safety core, (B) "general-purpose AI" sections, and (C) ad hoc elements—can make it easier to navigate. Most of our discussion in this Part II centers on the Act's product safety core. But we cover each of these three aspects in more detail.

A. THE CORE PRODUCT SAFETY REGIME: THE "NEW LEGISLATIVE FRAMEWORK"

The core of the AI Act, its regulation of high-risk AI systems, adopts the EU's framework approach to product safety.[57] This "New Legislative Framework" (NLF) has been laid out in two regulations (2008, 2019)[58] and one directive (2008),[59] and in product-specific laws.[60] The procedural core of the

54. *But see, e.g.*, Alessandro Mantelero, BEYOND DATA: HUMAN RIGHTS, ETHICAL AND SOCIAL IMPACT ASSESSMENT IN AI (2022) (arguing for transplanting international law's human rights impact assessment regime into data protection and AI law).

55. Claire Boine & David Rolnick, *Why the AI Act Fails to Understand Generative AI*, 26 MINN. J.L. SCI. & TECH. 61 (2025).

56. *See, e.g.*, Kaminski & Malgieri, *supra* note 36.

57. Michael Veale & Frederik Zuiderveen Borgesius, *Demystifying the Draft EU Artificial Intelligence Act*, 22 COMPUT. L. REV. INT'L 97 (2021); Nicolas Petit & Marco Almada, *The EU AI Act: Between the Rock of Product Safety and the Hard Place of Fundamental Rights,* 62 COMMON MKT. L. REV. 85 (2025).

58. Regulation (EC) No. 765/2008, 2008 O.J. (L 218) 30, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32008R0765; Regulation (EU) 2019/1020, 2019 O.J. (L 169) 1, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32019R1020.

59. Decision 768/2008/EC, 2008 O.J. (L 218) 82, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32008D0768.

60. *New Legislative Framework*, EUR. COMM'N, https://single-market-economy.ec.europa.eu/single-market/goods/new-legislative-framework_en.

AI Act thus adopts the same framework used for regulating products such as medical devices, construction products, and children's toys.[61]

The NLF operates roughly as follows: in order to legally be distributed on the EU market, providers of a product must undergo a "conformity assessment" that indicates the product and its safety governance are in compliance with EU law. After undergoing the conformity assessment, providers of the product may label it as "CE" and then circulate the product on the internal market. Labeled products are then subject to post-market supervision by entities known as "market surveillance authorities." We go through all of this specific to the AI Act in more detail below.

First, we discuss the Act's risk tiers in Section III.A.1. Then we briefly discuss its bans in Section III.A.2. We then, in much greater detail, discuss the Act's regulation of high-risk AI systems, which constitutes the bulk of the Act, and its use of the scaffolding of the New Legislative Framework in Section III.A.3.

1. The Risk Tiers

The Act structures its core framework around three tiers of risks raised by particular applications of AI: unacceptable risks, high risks, and everything else. It can be helpful to think of the tiers as a traffic light.[62] The Act prohibits a specific list of AI practices in Chapter II (red light).[63] It regulates a potentially evolving list of AI practices in Chapter III (yellow light).[64] And it permits everything else, subject to suggested self-regulation (green light).[65] The bulk of the Act focuses on the regulation of high-risk AI systems (yellow light). We do the same here.

We make one high-level observation before proceeding. First note that the Act regulates two primary sets of actors: AI providers and AI deployers.[66] (It also regulates AI distributors and importers, which makes sense if you think of the law as being concerned with the introduction of a product onto the EU

61. *Id.*

62. Nicolas Petit points out that the traffic light analogy only goes so far, as an AI system can actually fall into multiple categories at once. But we are trying to produce a simplified guide, so let's stick with the traffic light, for now.

63. AI Act, *supra* note 2, art. 5.

64. *Id.* ch. 3 (High-Risk AI Systems); *see also id.* Annex III.

65. *See, e.g.*, AI Act, *supra* note 2, art. 95(1) (suggesting voluntary Codes of Conduct) ("The AI Office and the Member States shall encourage and facilitate the drawing up of codes of conduct, including related governance mechanisms, intended to foster the voluntary application to AI systems, other than high-risk AI systems, of some or all of the requirements set out in Chapter III, Section 2 taking into account the available technical solutions and industry best practices allowing for the application of such requirements.").

66. *Id.* art. 2(1)(c).

market.[67]) The law's primary focus is on AI providers—that is, what many refer to as AI developers.[68] It also, however, regulates AI deployers: the users of AI systems.[69]

The core structure of the Act, then, regulates a particular mental model of AI: AI that is developed by one actor and used by another, *for a particular purpose*. The Act's primary obligations attempt to trace accountability as it passes between these two categories of actors.[70] Which risk tier applies to a particular "AI practice" depends on the specific purpose of the particular AI system. It should come as little surprise, then, that this purpose-centric framing breaks down when it comes to general-purpose AI systems.[71]

2. The Bans

The AI Act's regulation of the riskiest tier employs a simple but serious legal mechanism: the ban.

Although a short Chapter, the AI Act's regulation of Prohibited AI Practices through bans is in itself rather remarkable. Some uses of AI systems have been banned or paused in the United States, on the municipal level[72] or the state level.[73] There have also been discussions of banning certain uses of AI internationally, in war.[74] But generally speaking, as of the enactment of the

67. In the interest of relative brevity, we do not cover these obligations in any detail here. Suffice it to say that they consist largely of making sure the AI provider is in compliance with the law, including by checking for conformity markings. *Id.* art. 2(1)(d); *see also id.* arts. 23 (importers)–24 (distributors).

68. This also includes the entity having an AI system developed. *Id.* art. 3(3) ("natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge").

69. *Id.* art. 3(4) ("a natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity").

70. *See id.* art. 13 (discussing transparency to deployers); *see also id.* art. 14 (enabling human oversight); *see also* Rebecca Crootof, Margot E. Kaminski & W. Nicholson Price II, *Humans in the Loop*, 76 VAND. L. REV. 429, 503 (2023) (discussing the draft AI Act's approach to human oversight). The Act also creates some specific requirements for importers and distributors, but those are designed primarily to prevent loopholes—for example, creating a system abroad without complying. AI Act, *supra* note 2, arts. 23–24.

71. Boine & Rolnick, *supra* note 55.

72. *See* AP, *Seattle Mayor Ends Police Drone Efforts*, USA TODAY, https://www.usatoday.com/story/news/nation/2013/02/07/seattle-police-drone-efforts/1900785/ (last updated Feb. 7, 2013).

73. S.B. 25-143, Gen. Assemb., Reg. Sess. (Colo. 2025).

74. STOP KILLER ROBOTS, https://www.stopkillerrobots.org/; *see* Rebecca Crootof, *The Killer Robots Are Here: Legal and Policy Implications*, 36 CARDOZO L. REV. 1837 (2015) (discussing bans versus regulation).

AI Act, few AI systems have been meaningfully regulated, let alone banned in the United States.

As a matter of regulatory design, the AI Act's use of bans sets a relatively clear set of rules for certain specific AI practices. If regulated companies truly want clarity, a ban on certain practices offers it. A ban is not a standard.[75] It does not allow (much) room for interpretation in application.[76] Nor does it delegate decision-making to other bodies, like agencies or courts. The decision of what to ban is made by public lawmakers, not some set of private actors operating through technical standards-setting institutions.[77] And although much of what the AI Act does is lighter-touch regulation, this component most certainly is not. Recall that the stated purpose of the AI Act is to facilitate the uptake of AI systems on the EU market. The Act's bans of certain uses of AI systems in Chapter II act as an ostensible backstop to that overarching goal of uptake through regulation, offering at least symbolic outer limits to what Europe will allow.[78]

The AI Act bans a list of "Prohibited AI Practices" in the text of the Act itself, rather than an annex.[79] This makes the list stickier and harder to amend than, say, the Act's list of "high-risk" practices.[80] It ostensibly reflects a European consensus on which specific uses of AI de facto violate fundamental rights or pose too high a threat to health or safety.[81]

---

75. Louis Kaplow, *Rules v. Standards: An Economic Analysis*, 42 DUKE L.J. 557 (1992). *But see* Pierre Schlag, RULES AND STANDARDS, 33 UCLA L. REV. 379 (1985).

76. There are exceptions, *see, e.g.*, the exception for uses of emotional inference AI systems in the workplace and educational settings for medical or safety reasons. AI Act, *supra* note 2, art. 5(1)(f).

77. See *id.* arts. 112(1), 112(11) for how amendments to the list of banned practices might happen.

78. *See generally id.* ch. 2.

79. *Id.* art. 5.

80. The default is that the AI Act may be amended through the ordinary EU legislative process, which involves all three of the Commission, Parliament, and Council of the European Union. This takes longer than the streamlined process envisioned, for example, for updating Annex III of the AI Act, discussed below. *See* AI Act, *supra* note 2, art. 112(1) ("The Commission shall assess the need for amendment of the list set out in Annex III and of the list of prohibited AI practices laid down in Article 5, once a year following the entry into force of this Regulation, and until the end of the period of the delegation of power laid down in Article 97. The Commission shall submit the findings of that assessment to the European Parliament and the Council."). Art. 112(11) envisions the AI Office in the Commission as taking a lead role in coming up with "an objective and participative methodology for the evaluation of risk levels based on the criteria outlined in the relevant Articles and the inclusion of new systems in: . . . (b) the list of prohibited practices set out in Article 5 . . . ." *Id.*

81. *See* Ljupcho Grozdanovski & Jérôme De Cooman, *Forget the Facts, Aim for the Rights! On the Obsolescence of Empirical Knowledge in Defining the Risk/Rights-Based Approach to AI Regulation in the European Union*, 49 RUTGERS COMPUT. & TECH. L.J. 207 (2023) (for criticism of the arbitrariness of the list).

In reality, however, the list appears to be a response to a number of salient news stories about the misuse of algorithms or AI systems. Banned practices include: the use of AI for social scoring (the "let's-not-be China ban");[82] the use of AI for predicting the risk of committing a criminal offense (the "Minority Report ban");[83] the use of AI to infer emotions at work or in an educational setting (the "Emotional Phrenology ban"[84]);[85] and the scraping of public images to contribute to the creation of a facial recognition system (the "Clearview AI ban")[86] The full list is outlined in Chapter II, Article 5.[87]

One prohibited practice that merits further discussion is the ban on the use of real-time biometric systems, in public spaces, for law enforcement use (we call it "the biometrics ban" for the rest of this subpart).[88] That is for two reasons. First, the biometrics ban establishes a ban in close proximity to uses permitted but regulated by other portions of the Act. The Act bans only a very specific subset of uses of biometric systems: real-time (as opposed to after-the-fact), in public spaces (as opposed to in private spaces), and for law enforcement use (as opposed to private sector use).[89] Otherwise, biometric

82. AI Act, *supra* note 2, art. 5(1)(c).

83. *Id.* art. 5(1)(d).

84. Luke Stark & Jevan Hutson, *Physiognomic Artificial Intelligence*, 32 FORDHAM INTELL. PROP., MEDIA & ENT. L.J. 922 (2021) (calling it phrenology/physiognomy).

85. AI Act, *supra* note 2, art. 5(1)(f) ("the placing on the market, the putting into service for this specific purpose, or the use of AI systems to infer emotions of a natural person in the areas of workplace and education institutions, except where the use of the AI system is intended to be put in place or into the market for medical or safety reasons").

86. *Id.* art. 5(1)(e).

87. In addition to the named exceptions, *supra* note 76, it includes: the use of subliminal techniques, *id.* art. 5(1)(a), materially distorting persons' behavior through exploiting known vulnerabilities, *id.* art. 5(1)(b), and the making of certain inferences (race, political opinion) through biometric data, *id.* art. 5(1)(g).

88. *Id.* art. 5(1)(h).

89. *See* Veale & Zuiderveen Borgesius, *supra* note 57 ("only 'real-time' systems that capture, compare, and identify 'instantaneously, near-instantaneously or in any event without a significant delay' are prohibited. This excludes 'post' systems which, for example, biometrically analyse footage after an event, for example to identify individuals at protests after-the-fact, and systems that categorise individuals biometrically. As online spaces are also out-of-scope, live biometric identification on e.g., video streams is also excluded."); *see also Cop Out: Security Exemptions in the Artificial Intelligence Act*, STATEWATCH, https://www.statewatch.org/automating-authority-artificial-intelligence-in-european-police-and-border-regimes/2-cop-out-security-exemptions-in-the-artificial-intelligence-act/ (referring to "(Un)prohibited practices" and observing that "despite supposed bans on practices such as profiling, biometric categorization, and mass biometric surveillance, law enforcement and migration authorities enjoy numerous exemptions that may enable widespread deployment of these techniques"); Francesca Palmiotto, *The AI Act Roller Coaster: The Evolution of Fundamental Rights Protection in the Legislative Process and the Future of the Regulation*, 16 EURO. J. RISK REG. 770, 780, 789 (2025) ("The list of prohibited AI practices in Article 5 has been the most debated

systems are regulated under the Act as high-risk uses of AI.[90] This establishes a potential cliff effect, whereby regulated AI providers may do their darndest not to fall into the prohibited category of behavior. Some have noted, too, that the Act does not ban but rather condones the installation of the infrastructure for facial recognition.[91]

Second, the biometrics ban, which was much-debated during drafting, is the only ban—and indeed, one of the few places in the Act—to envision and structure significant interplay with courts. Unlike several of the other bans, the biometrics ban contains three named exceptions: for a targeted search for victims, for a "specific, substantial and imminent threat to the life or physical safety" or risk of a terrorist attack, and for finding or identifying a person suspected of having committed certain more serious crimes.[92] It sets up a system of prior judicial oversight (or similar independent administrative oversight, depending on national law) over these exceptions.[93] This is notable because it invokes courts directly as the authority on fundamental rights oversight, unlike other aspects of the Act.

3. High-Risk AI Systems

Most of the AI Act concerns the regulation of high-risk AI systems—the yellow in our traffic light of risk tiers.[94] Here, we discuss: which AI systems are considered "high risk," the substance of the regulation for each of providers

---

of the AI Act, particularly regarding real-time biometric identification . . . . The AI Act embeds double standards for individuals affected by AI systems, with lower protection for individuals suspected or accused of having committed a crime, migrants, asylum seekers and refugees. From a legal and ethical perspective, this is a critical weakness of the Regulation.").

90. *See* AI Act, *supra* note 2, Annex III (defining high-risk AI systems to include):

> Biometrics, in so far as their use is permitted under relevant Union or national law:
>
> (a) remote biometric identification systems. This shall not include AI systems intended to be used for biometric verification the sole purpose of which is to confirm that a specific natural person is the person he or she claims to be;
>
> (b) AI systems intended to be used for biometric categorisation, according to sensitive or protected attributes or characteristics based on the inference of those attributes or characteristics;
>
> (c) AI systems intended to be used for emotion recognition.

91. Veale & Zuiderveen Borgesius, *supra* note 57, at 102 ("any authorisation of biometrics necessitates installing re-purposable infrastructure. Many already argue the Draft AI Act legitimises rather than prohibits population-scale surveillance").

92. AI Act, ch. 2, art. 5(1)(h)(i)–(iii); *see also id.* Annex II (listing covered criminal offenses).

93. *Id.* art. 5(3).

94. AI Act, *supra* note 2, ch. III.

(developers) and deployers (users) of high-risk AI systems, how conformity assessment works specific to the Act (including the central role of private technical standards-setting), and the complex accountability systems established by the Act (including registration and post-market monitoring).

a) What Practices Are "High Risk"?

An AI system is considered "high risk" if it is a product or safety component of a product covered by certain named EU laws under the New Legislative Framework[95] and elsewhere.[96] Additionally, an AI system is considered "high risk" if it is named in Annex III of the Act.[97] Annex III currently lists eight categories of high-risk AI systems. These categories include: biometric systems that weren't banned in the Act and are otherwise permitted under Union and national law;[98] AI systems used in critical infrastructure;[99] certain AI systems used in educational and vocational training;[100] certain AI systems used in employment, including in recruitment, termination, and employee monitoring;[101] AI systems used for determining certain public and private essential benefits;[102] and three more.[103]

Aiming to future-proof the law, the AI Act makes the list of high-risk AI systems in Annex III easier to update than typical EU law. Article 7 gives the Commission the authority to adopt "delegated acts" to update Annex III, so long as the risk of harm is as high as existing practices on the current list and the new high-risk systems fall under the same existing eight categories.[104]

There are two wrinkles of note. First, some of the practices listed in Annex III may in practice be banned, if their use is not permitted by other EU or national law.[105] That is, developers should not assume that an AI practice

95. AI Act, *supra* note 2, art. 6(1). See also *id.* Annex I(A) for the list of relevant EU laws. That list includes: machinery, toys, recreation craft and personal watercraft, elevators, medical devices, and more.

96. See *id.* Annex I(B), which includes railroads, aircraft, and quadricycles, among other things.

97. *Id.* art. 6(2); *see also id.* Annex III.

98. *Id.* Annex III(1).

99. *Id.* Annex III(2).

100. *Id.* Annex III(3).

101. *Id.* Annex III(4).

102. *Id.* Annex III(5).

103. The remaining three are: certain law enforcement uses (Annex III(6)); certain uses in managing asylum, migration, and border control (Annex III(7)); and uses by judges (Annex III(8)(a)) and to influence elections (Annex III(8)(b)).

104. AI Act, *supra* note 2, art. 7.

105. *See, e.g.*, *id.* Annex III(6)–(7) ("in so far as their use is permitted under relevant Union or national law").

named in Annex III is necessarily legal; rather, it is just not banned by the AI Act specifically.[106]

Second, Article 6 contains a significant potential loophole with respect to the Act's high-risk categorization.[107] Providers can, in effect, opt out of having their systems designated as high-risk in some circumstances. Largely, these circumstances involve using AI as a de minimis element of an otherwise human decision-making process, even where that decision-making process falls into one of the eight categories in Annex III.[108] This creates incentives to put more humans in the loop for these kinds of decisions, whether authentically or performatively, in order to escape the Act's requirements.[109]

A provider may themselves determine under those circumstances that an AI system "does not pose a significant risk of harm to the health, safety or fundamental rights of natural persons, including by not materially influencing the outcome of decision making."[110] There is no ex ante scrutiny of this decision. The provider must document this determination and register the

106. *See also* European Data Protection Board-European Data Protection Supervisor, Joint Opinion 5/2021 on the Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) (June 18, 2021), https://www.edpb.europa.eu/our-work-tools/our-documents/edpbedps-joint-opinion/edpb-edps-joint-opinion-52021-proposal_en [hereinafter EDPB-EDPS Joint Opinion].

107. *See* AI Act, *supra* note 2, art. 6(3). The AI Act again establishes an easier lawmaking process by allowing for amendment through delegated acts "by adding new conditions to those laid down therein, or by modifying them, where there is concrete and reliable evidence of the existence of AI systems that fall under the scope of Annex III, but do not pose a significant risk of harm to the health, safety or fundamental rights of natural persons." *Id.* art. 6(6).

108. *Id.* art. 6(3) states that the exception "shall apply where *any* of the following conditions is fulfilled:

- (a) the AI system is intended to perform a narrow procedural task;
- (b) the AI system is intended to improve the result of a previously completed human activity;
- (c) the AI system is intended to detect decision-making patterns or deviations from prior decision-making patterns and is not meant to replace or influence the previously completed human assessment, without proper human review; or
- (d) the AI system is intended to perform a preparatory task to an assessment relevant for the purposes of the use cases listed in Annex III."

(emphasis added).

109. *See* Crootof et al., *supra* note 70, at 450 ("Some laws encourage regulatory arbitrage when retaining a human in the loop allows a system's developers or users to avoid more onerous regulation.").

110. AI Act, *supra* note 2, art. 6(3). However, a provider cannot self-designate as not-high-risk an AI system that "performs profiling of natural persons."

system before putting it on the market, so that the determination can be examined after the fact if harms should occur or contrary evidence should arise.[111]

b) What Are the Substantive Requirements for Providers (Developers) of High-Risk AI Systems?

The bulk of the AI Act aims to regulate providers of high-risk AI systems. Most of these requirements are outlined in Chapter III, Section 2 (helpfully titled "Requirements for high-risk AI systems").[112] Requirements include, among other things: establishing a risk mitigation system;[113] establishing data governance;[114] establishing technical documentation[115] and automated logging;[116] and establishing certain accuracy, robustness, and cybersecurity measures.[117]

Some of these requirements are more substantive. Others primarily establish procedures and documentation towards enabling later external accountability. Some entwine procedure with substance. And several are aimed at maintaining accountability during a handoff from system developer to system deployer.[118]

This section, like much of this Article, is not intended to provide legal advice nor to be exhaustive. We provide an overview of the Act's requirements that apply to providers of high-risk AI systems, with several more in-depth examples of each kind of requirement.

First, the AI Act, for all its procedural and accountability scaffolding, does contain substantive requirements. There are several examples in Article 15, on "Accuracy, robustness and cybersecurity."[119] The Act, for example, requires high-risk AI systems to "achieve an appropriate level of accuracy,"[120] and refers to developing substantive "benchmarks and measurement methodologies" as performance metrics.[121] The Act requires that developers disclose said levels of accuracy and metrics in instructions conveyed to deployers.[122]

111. *Id.* art. 6(4). National competent authorities may later ask for this documentation.
112. *Id.* ch. III, sec. 2.
113. *Id.* art. 9.
114. *Id.* art. 10.
115. *Id.* art. 11.
116. *Id.* arts. 12, 19 (obligation to retain logs).
117. *Id.* art. 15.
118. *See, e.g.*, *id.* art. 14 (on human oversight), *id.* art. 13 (on transparency to deployers).
119. *Id.* art. 15.
120. *Id.* art. 15(1).
121. *Id.* art. 15(2).
122. *Id.* art. 15(3).

Other examples of substantive requirements can be found in Article 10 on data and data governance.[123] The Act requires that "[t]raining, validation and testing data sets shall be relevant, sufficiently representative, and to the best extent possible, free of errors and complete in view of the intended purpose."[124] It additionally requires that such data sets must "have the appropriate statistical properties . . . as regards the persons or groups of persons in relation to whom the high-risk AI system is intended to be used."[125] The latter requirement is significant, given high-profile reports of algorithms developed using data from one population but then used on another, without confirming that the two populations had the same statistical properties.[126]

Second, elements of this Section of the Act require documentation and process towards establishing external accountability. For example, Article 11 requires technical documentation.[127] The elements of such technical documentation are outlined in Annex IV, and smaller businesses may adopt a simplified version.[128] The explicit purpose of the requirement of technical documentation is to establish external accountability to government actors.[129]

Third, elements of this Section of the Act involve process mixed with substance. For example, Article 9 requires that developers establish a "risk management system" for high-risk AI systems.[130] At first glance, these requirements look primarily procedural, requiring a set of steps.[131] But the risk

123. *Id.* art. 10.

124. *Id.* art. 10(3).

125. *Id.* arts. 10(3)–(4) ("Data sets shall take into account, to the extent required by the intended purpose, the characteristics or elements that are particular to the specific geographical, contextual, behavioural or functional setting within which the high-risk AI system is intended to be used.").

126. *See* State v. Loomis, 2016 WI 68, ¶¶ 63–66 (noting that COMPAS, which was trained on data from Broward County, Florida, and applied in Wisconsin, had not yet had a cross-validation study for the Wisconsin population). For more examples, see Angelina Wang, Sayash Kapoor, Solon Barocas & Arvind Narayanan, *Against Predictive Optimization: On the Legitimacy of Decision-Making Algorithms That Optimize Predictive Accuracy*, 1 ACM J. RESPONSIBLE COMPUTING 9:1, 9:12–13 (Mar. 20, 2024) (discussing examples of "distribution shifts").

127. AI Act, *supra* note 2, art. 11.

128. *Id.* art. 11(1); *see also id.* Annex IV.

129. *Id.* art. 11(1) ("demonstrate that the high-risk AI system complies with the requirements set out in this Section and to provide national competent authorities and notified bodies with the necessary information in a clear and comprehensive form to assess the compliance of the AI system with those requirements.").

130. *Id.* art. 9(1) ("A risk management system shall be established, implemented, documented and maintained in relation to high-risk AI systems.").

131. Article 9(2) of the AI Act states that:

> (a) the identification and analysis of the known and the reasonably foreseeable risks that the high-risk AI system can pose to health, safety or

management system also has substantive aspects. For example, it requires mitigation of risks down to an "acceptable" level of risk, both for each hazard and for the system overall.[132] Article 9 also requires that high-risk AI systems be tested.[133] Substantively, it requires these systems be fit for their intended purpose.[134]

Fourth, several of the Act's requirements for developers aim to establish an accountability hand-off from providers (developers) to deployers (users) of high-risk AI systems. Article 13 requires that system developers create and transmit instructions for and transparency to deployers.[135] Instructions must include at least "the characteristics, capabilities and limitations of performance" of the AI system, including its purpose; its level of accuracy, robustness, and cybersecurity; circumstances that may lead to higher risks; and "its performance regarding specific persons or groups of persons on which the system is intended to be used."[136] The Act also requires explanations of the AI systems to deployers (as opposed to affected individuals, which is addressed in Article 86), for example requiring that providers disclose "where applicable, the technical capabilities and characteristics of the high-risk AI system to provide information that is relevant to explain its output" to deployers.[137]

In some places, however, the envisioned accountability hand-off may break down. Take, for example, the provisions on human oversight, which

> fundamental rights when the high-risk AI system is used in accordance with its intended purpose;
>
> (b) the estimation and evaluation of the risks that may emerge when the high-risk AI system is used in accordance with its intended purpose, and under conditions of reasonably foreseeable misuse;
>
> (c) the evaluation of other risks possibly arising, based on the analysis of data gathered from the post-market monitoring system referred to in Article 72;
>
> (d) the adoption of appropriate and targeted risk management measures designed to address the risks identified pursuant to point (a).

132. *Id.* art. 9(5).

133. *Id.* art. 9(6) ("High-risk AI systems shall be tested for the purpose of identifying the most appropriate and targeted risk management measures. Testing shall ensure that high-risk AI systems perform consistently for their intended purpose and that they are in compliance with the requirements set out in this Section.").

134. *Id.* ("Testing shall ensure that high-risk AI systems perform consistently for their intended purpose").

135. *Id.* arts. 13(1)–(2).

136. *Id.* art. 13(3).

137. *Compare id.* art. 13(3)(b)(vii) ("where applicable, information to enable deployers to interpret the output of the high-risk AI system and use it appropriately"), *with* GDPR, *supra* note 19, art. 15(h) (requiring disclosure to affected individuals of "meaningful information about the logic involved"); *see also* Case C-203/22, *supra* note 36 (interpreting art. 15 of GDPR in light of art. 22).

describe a complex pass-off from providers to deployers. The Act requires that high-risk AI systems be designed by providers for human oversight "commensurate with the risks, level of autonomy and context of use."[138] AI providers are responsible for building related enabling features into AI.[139] Providers are also responsible for telling deployers how to use high-risk AI, including how much human oversight is needed.[140] However, the Act recognizes that human oversight itself will often be implemented by deployers.[141] Thus, the envisioned pass-off ideally goes as follows: AI providers design their systems for human oversight (as a form of risk mitigation) and instruct deployers as to how much oversight should be used. Then, deployers must follow the instructions.

The Act however, largely declines to put direct obligations on the deployers.[142] As written, the Act delegates to AI providers to decide what level of human oversight is required, and to govern deployers through instructions for use.[143] Michael Veale and Frederik Zuiderveen Borgesius describe the set-up as follows: "Somewhat strangely, no obligations for human oversight flow directly from the Act to a user. In relation to human oversight, users must simply follow the instruction manual."[144] The only direct obligations on deployers related to human oversight are to assign it to "natural persons who have the necessary competence, training and authority, as well as the necessary support,"[145] and to use high-risk AI systems "in accordance with the

138. AI Act, *supra* note 2, art. 14(3).

139. *Id.* art. 14(3)(a).

140. *Id.* art. 13(3)(d) ("the human oversight measures referred to in Article 14, including the technical measures put in place to facilitate the interpretation of the outputs of the high-risk AI systems by the deployers").

141. *Id.* art. 14(3)(b) ("measures identified by the provider before placing the high-risk AI system on the market or putting it into service and that are appropriate to be implemented by the deployer").

142. *See* Veale & Zuiderveen Borgesius, *supra* note 57, at 104 ("Somewhat strangely, no obligations for human oversight flow directly from the Act to a user. In relation to human oversight, users must simply follow the instruction manual."); *see also* Crootof et al., *supra* note 70, at 503–04 ("The Act divides regulated entities into providers, who build AI systems, and users, who use them. As a consequence, *nobody is really responsible for the human-machine system as a whole*.").

143. There is one notable exception: the AI Act requires that a deployer may not take action on the basis of identification by remote biometric identification systems unless at least two humans "with the necessary competence, training and authority" have separately confirmed the identification. AI Act, *supra* note 2, art. 14(5). The Act exempts, however, "high-risk AI systems used for the purposes of law enforcement, migration, border control or asylum, where Union or national law considers the application of this requirement to be disproportionate." *Id.*

144. Veale & Zuiderveen Borgesius, *supra* note 57, at 104.

145. AI Act, *supra* note 2, art. 26(2).

instructions for use."[146] Some—but not all—deployers must provide a "description of the implementation of human oversight measures, according to the instructions for use."[147]

c) What Are The Substantive Requirements for Deployers (Users) of High-Risk AI Systems?

This leads us into a much briefer discussion of the Act's requirements for deployers (users) of high-risk AI systems. Largely, these are laid out in Article 26.[148] But deployers should be sure to read the Act as a whole, as several requirements for deployers are laid out elsewhere.[149] We outline a few core requirements here.

As noted above, deployers are legally obligated to follow the instructions for use that they receive from providers, including any instructions regarding human oversight.[150] They are required to ensure that input data is "relevant and sufficiently representative in view of the intended purpose of the high-risk AI system."[151] They are required to keep logs automatically generated by AI systems, typically for at least six months.[152] And, centrally, deployers are required to "monitor the operation of the high-risk AI system" and notify the provider, distributor, and relevant government authority when things go wrong and the system presents a risk even when instructions are followed, or when there has been a "serious incident."[153] They are required in those cases to stop using the AI system.[154]

Deployers are also assigned several notification requirements. They must notify workers and their representatives before deploying high-risk AI systems in the workplace.[155] And they must notify affected individuals if they have been subject to a decision made by a high-risk AI system.[156] Deployers are also subject to several requirements outlined elsewhere in the Act, which we discuss further below: to provide individual explanations of AI decisions,[157] and to

146. *Id.* art. 26(1).

147. Only if they are required to conduct a Fundamental Rights Impact Assessment. *See id.* art. 27(1)(e).

148. *Id.* art. 26.

149. *See, e.g.*, our discussion of "ad hoc" elements below, and discussion of biometric verification in note 143 .

150. AI Act, *supra* note 2, art. 26(1).

151. *Id.* art. 26(4).

152. *Id.* art. 26(6).

153. *Id.* art. 26(5).

154. *Id.*

155. *Id.* art. 26(7).

156. *Id.* art. 26(11).

157. *Id.* art. 86

conduct Fundamental Rights Impact Assessments in some contexts, before deploying a system.[158]

d) Conformity Assessments

As mentioned, the key feature of the NLF that the AI Act borrows is the "conformity assessment" and the accountability framework that accompanies it.[159] A conformity assessment is a document that states whether the "essential requirements" of the law in question are met, as well as different procedural requirements that the law may implement.[160] In the AI Act, a conformity assessment requires that three categories of requirements be satisfied.[161]

First is the quality management system. Article 17 lays out thirteen requirements for the quality management system, among them "a strategy for regulatory compliance"; "techniques, procedures and systematic actions" for design, design control, design verification, development, quality control, and quality assurance; "examination, test and validation procedures to be carried out before, during and after the development of the high-risk AI system"; "technical specifications, including standards, to be applied"; comprehensive "systems and procedures for data management"; "the risk management system referred to in Article 9"; "the setting-up, implementation and maintenance of a post-market monitoring system"; "procedures related to the reporting of a serious incident"; and "an accountability framework setting out the responsibilities of the management and other staff with regard to all the aspects listed."[162]

The second requirement that must be satisfied is that technical documentation must demonstrate compliance with the essential requirements in Chapter III, Section 2 of the law.[163] Chapter III, Section 2 of the law corresponds to Articles 8–15, so this entails the risk management framework, data governance system, documentation, logging, instructions for use, human oversight, and accuracy and robustness checks described above.[164]

The third requirement that must be satisfied is verification of a post-market monitoring system described in Article 72. This is a system of monitoring for data that allows a provider to evaluate continuous compliance

158. *Id.* art. 27.
159. *Id.* art. 43.
160. *Id.* art. 3(20).
161. *Id.* Annexes VI–VII.
162. *Id.* art. 17(1).
163. *Id.* art. 3(20).
164. *Id.* arts. 9–15.

with the regulation throughout the lifetime of the AI system.[165] The Commission is charged with establishing a template for this by early 2026.[166]

There are two sets of procedures for undergoing a conformity assessment. One is a self-assessment laid out in Annex VI, and the other is a third-party assessment described in Annex VII.[167] Article 43 dictates which procedures each type of high-risk AI system is subject to. It divides high-risk AI systems into three categories with respect to whether they (a) can opt for either self-certification or third-party assessment, (b) may use the self-certification procedure, or (c) must use third-party assessment.

The bottom line is that (1) providers of biometric systems that follow published standards can self-certify, but if they do not, they must get third-party certification;[168] (2) providers of AI for critical infrastructure and other fundamental-rights-impacting AI can self-certify;[169] and (3) providers of safety-impacting AI—which is safety-impacting only because it is one of the types of products *otherwise* designated as safety-impacting under the NLF—must undergo the third-party assessment procedure dictated by the other product safety law the product is subject to.[170] The fact that rights-impacting AI is governed by self-certification of compliance has led to some of the serious critiques by European scholars and activists that this law does not take fundamental rights seriously enough.[171]

---

165. *Id.* art. 72(2).

166. *Id.* art. 72(3).

167. *Id.* art. 43.

168. *Id.* art. 43(1). Under Article 43(1), providers of "high-risk AI systems listed in point 1 of Annex III,"—otherwise known as the biometric systems that are not banned outright—can opt for self-certification or third-party assessment if they've followed harmonized standards under Article 40 or the common procedures under Article 41 (more on both below). However, if such standards do not exist or were not followed, the provider must use the Annex VII (third-party assessment) procedure.

169. *Id.* art. 43(2). Article 43(2) says that providers of "high-risk AI systems referred to in points 2 to 8 of Annex III"—otherwise known as the non-biometric rights-impacting AI (points 3–8) plus critical infrastructure (point 2)—may use the Annex VI (self-certification) procedure.

170. *Id.* art. 43(3). Article 43(3) says that providers of "high-risk AI systems covered by the Union harmonisation legislation listed in Section A of Annex I"—otherwise known as safety-impacting AI systems already covered by a different NLF law—should follow the conformity assessment procedures of those other laws and just add the requirements of the AI Act to those conformity assessments.

171. *See* Daniel Leufer, Fanny Hidvegi & Alessia Zornetta, *The Pitfalls of the European Union's Risk-Based Approach to Digital Rulemaking*, 71 UCLA L. REV. DISCOURSE 156 (2024). We would add that while a cynic—or an American—might not be entirely surprised by fundamental rights being pushed aside, we find the idea that self-certification also applies to critical infrastructure in some ways even more surprising, leading us to wonder whether there is a more fundamental respect for law in play in Europe, such that we should not assume law

While the Act contains these requirements for conformity assessments, in practice many of the details will end up being developed by private standards-setting organizations. Article 40 directs the Commission to issue standardization requests to the ESOs to cover all substantive requirements of the law. Article 40(1) then creates a kind of safe harbor: it states that "[h]igh-risk AI systems or general-purpose AI models which are in conformity with harmonised standards or parts thereof . . . shall be presumed to be in conformity" with the requirements of the law.[172]

Thus, providers can avoid the sometimes vague and difficult questions about when and whether their process lines up with the requirements of the law by adhering strictly to any published standards. The Commission has requested that a joint technical committee of CEN and CENELEC take up the task, which is not yet completed.[173] Assuming they finish, and the standards meet the criteria in the law,[174] then those standards are poised to become the primary way that companies comply with the AI Act.[175]

e) Accountability and Enforcement

The AI Act's envisioned accountability framework for high-risk systems is based on (a) establishing internal corporate governance, (b) creating documentation that overseeing institutions can examine ex post, (c) setting up systems that will alert overseeing institutions if things go wrong, and (d) enabling market surveillance authorities to obtain access to information and affording them the ability to take remedial or punitive action.

The first component, corporate governance, is accomplished via the Chapter III, Section 2 articles described in the last two parts—for example, creation of risk management systems (Art. 9); data governance systems (Art. 10); human oversight (Art. 14); and checks on accuracy, robustness, and security (Art. 15).[176] The conformity assessment, whether self-certified or

must be written with Holmes' bad man in mind. On the other hand, the EDPB and EDPS called for third-party oversight and did not succeed in getting it into the law. *See* EDPB-EDPS Joint Opinion, *supra* note 106.

172. AI Act, *supra* note 2, art. 40(1).

173. *Artificial Intelligence*, *supra* note 24; *Commission Implementing Decision on a Standardisation Request to the European Committee for Standardisation and the European Committee for Electrotechnical Standardisation in support of Union Policy on Artificial Intelligence*, at 5, C(2023) 3215 final (May 22, 2023), https://ec.europa.eu/transparency/documents-register/detail?ref=C(2023)3215&lang=en.

174. Shortcomings in either the standards or common specifications are a reason that market surveillance authorities may reject the presumption. AI Act, *supra* note 2, art. 79(6).

175. If the standards are unfinished or inadequate, under Article 41, the Commission may adopt "common specifications" that functionally take the place of technical standards. *See, e.g.*, *id.* art. 43(1) (referring to both Articles 40 and 41 as alternatives to each other).

176. *See* Sections II.D(2)–(3).

certified by a notified body, requires an assessment that these systems have been adequately set up. Similarly, the conformity assessment requires a statement of compliance with Article 17, which requires a "quality management system in place that ensures compliance."[177]

The Act's second component of accountability and enforcement is documentation. Article 11 requires technical documentation to be kept up to date throughout the life of the product.[178] Article 17 requires that the quality management system "be documented in a systematic and orderly manner in the form of written policies, procedures and instructions."[179] Both are necessary components of the corporate governance framework. But the real point of the documentation is enabling later oversight by government bodies. Article 18 requires the technical and quality management system documentation be kept for ten years "at the disposal of the national competent authorities,"[180] and Article 21 requires that providers give a requesting authority "all the information and documentation necessary to demonstrate the conformity of the high-risk AI system . . . ."[181] Article 74 grants "full access by providers to the documentation as well as the training, validation and testing data sets used for the development of high-risk AI systems."[182] Similarly, Article 77 grants national authorities overseeing fundamental rights access to "any documentation created or maintained under this Regulation in accessible language and format when access to that documentation is necessary for effectively fulfilling their mandates within the limits of their jurisdiction."[183]

The third component of the Act's system of accountability and enforcement is affirmative notice to government actors. Article 20 begins with self-governance, requiring that whenever a provider learns that a system is no longer in conformity, they must take immediate corrective action "to bring that system into conformity, to withdraw it, to disable it, or to recall it, as appropriate."[184] But importantly, it also states that where a system "present[s] risks to the health or safety, or to fundamental rights, of persons"[185] and the provider becomes aware of the risk, it must investigate and, along with the

177. AI Act, *supra* note 2, art. 17(1)(a).
178. *Id.* art. 11(1).
179. *Id.* art. 17(1).
180. *Id.* art. 18(1).
181. *Id.* art. 21(1).
182. *Id.* art. 74(12). A similar provision in Article 91 grants the Commission access to documentation of general-purpose models, because the Commission is designated as the market surveillance authority for general-purpose AI. *See id.* arts. 91(1), 75(1); *see also infra* Section III.C.
183. AI Act, *supra* note 2, art. 77(1).
184. *Id.* art. 20(1).
185. *Id.* art. 79(1) (cited by art. 20(2)).

deployer of the system, inform the market surveillance authorities and notified body (where applicable) of the risk.[186] Combined with the requirements for automatic logging, this provision is designed to alert the government when anything goes wrong.

Finally, the fourth component of the Act's accountability framework is the ability of relevant government bodies to conduct investigations and take remedial and punitive action where necessary. This constitutes a framework of responsive regulation that is common among EU regulators.[187]

Article 79 instructs market surveillance authorities to evaluate systems presenting a risk to health, safety, or to fundamental rights once they are made aware of such a risk, and, if appropriate, to cooperate with national authorities that protect fundamental rights.[188] If the authority finds noncompliance, it must require the "relevant operator to take all appropriate corrective actions to bring the AI system into compliance, to withdraw the AI system from the market, or to recall it."[189] If the operator does not take the corrective action, the authority must escalate the matter, prohibiting the product or recalling it, as necessary.[190] Rules about "penalties and other enforcement measures, which may also include warnings and non-monetary measures," are delegated to member states.[191]

Article 99 provides for varied maximum fines depending on the violations. Violations of the bans in Article 5 are the highest (the higher of €35 million or 7 percent of annual revenue), with noncompliance with certain "provisions related to operators or notified bodies" (€15 million or 3 percent) in the middle, and probably the most common issue—supplying "incorrect, incomplete or misleading information to notified bodies or national competent authorities in reply to a request" (€7.5 million or 1 percent)—the lowest fine, but still substantial.[192]

Overall, this is a relatively light-touch approach to accountability and oversight, that gives primary control to the system providers and relies heavily on good-faith substantive compliance. The clearest example of this is the

186. *Id.* art. 20(2).

187. *See* William McGeveran, *Friending the Privacy Regulators*, 58 ARIZ. L. REV. 959, 983–85 (2016); Margot E. Kaminski, *Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability*, 92 S. CAL. L. REV. 1529, 1596–97 (2019); *see also* 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, 124 (2013) (on the generally collaborative approach of the pre-GDPR Dutch data protection regime).

188. AI Act, arts. 79(1)–(2).

189. *Id.* art. 79(2).

190. *Id.* art. 79(5).

191. *Id.* art. 99(1).

192. *Id.* arts. 99(3)–(5).

ability of a provider to designate a system as *not* high-risk despite its being a system on the list of high-risk types of systems.[193] The only requirements for opting the system out of the substantive regulations of this law almost entirely are to document the decision, to make that documentation available to the relevant national authorities upon request, and to register it in an EU database of high-risk systems.[194]

Other models of oversight are also notably absent. There is no private right of action for injured individuals; instead, the Act provides for a right to lodge a complaint with a market surveillance authority, which is then folded into the authority's standard process.[195] Transparency to the public, too, is limited. While there is extensive documentation, it is intended for government actors, not the public. The Act does provide for a public database of existing high-risk AI systems under Annex III (the rights-impacting plus critical infrastructure systems),[196] but nothing else.

The Act's accountability framework overall thus relies on the hope that AI providers know best how to judge their systems' risks, and that the government should largely monitor, getting involved only if and after things go wrong.[197] Again, this is not really a precautionary regulatory framework.

B. GENERAL-PURPOSE AI MODELS

The Act's approach to what it terms "general-purpose AI" models is distinct from the main portion of the law concerned with "high-risk AI" systems and works differently. It, too, adopts elements of risk regulation. But unlike the Act's approach to predictive AI, it is not built as squarely on the NLF.

The reason for this is in some sense a practical one: The original AI Act, including the reliance on the NLF, was drafted in 2021, before ChatGPT was released to the public and became a household name. When concerns over generative AI exploded in 2022, the law's drafters—who were writing a law

193. *See supra* notes 109–111 and accompanying text.

194. AI Act, *supra* note 2, art. 6(4). Recent proposals to revise the AI Act aim to get rid of this registration condition. *See* Maria Niestadt (Eur. Parl. Rsch. Serv.), *Briefing: Digital Omnibus on AI*, 3, PE 782.651 (Feb. 2026), https://www.europarl.europa.eu/RegData/etudes/BRIE/2026/782651/EPRS_BRI(2026)782651_EN.pdf.

195. *Id.* art. 85.

196. *Id.* art. 71.

197. *See* Veale & Zuiderveen Borgesius, *supra* note 57, at 102 ("The philosophy of the NLF is that '[t]he manufacturer, having detailed knowledge of the design and production process, is best placed to carry out the complete conformity assessment procedure. Conformity assessment should therefore remain the obligation of the manufacturer alone.' This distinguishes NLF regimes (including the Draft AI Act) from pharmaceutical regulation, where a public authority (e.g., the European Medicines Agency) carries out an assessment themselves before granting pre-marketing approval.").

purporting to cover all of AI—were forced to go back and add in some treatment of generative, or what they call "general-purpose," AI.[198]

Rather than rewrite the bill from the ground up, the drafters inserted new provisions that mimicked some of the concepts in the rest of the bill but used different enforcement mechanisms and institutions. Probably the biggest overall difference is that the Act treats these general-purpose AI models as inherently transnational, eschewing reliance on national market surveillance authorities and centralizing oversight at the EU level.

The interaction of these provisions with the rest of the Act is complicated. A general-purpose AI model might end up integrated into a general-purpose AI system, which are not assigned to any of the three risk tiers that make up the rest of the Act.[199] Or, it might end up integrated into an AI system that does fall into one of the Act's risk tiers.[200] In any event, the Act's general-purpose AI obligations outlined below fall on the providers of general-purpose AI models regardless of whether the system into which the model is integrated fits neatly into the rest of the regulation.

The Act contrasts generative AI models with other types of AI models based primarily on the generality of application, defining a "general-purpose AI model" as a model "that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications."[201] The definition "includ[es] where such an AI model is trained with a large amount of data using self-supervision at scale," but does not limit the definition to such technical parameters. It

198. Council of the EU Press Release 1008/22, Artificial Intelligence Act: Council Calls for Promoting Safe AI that Respects Fundamental Rights (Dec. 6, 2022) (noting new provisions about general-purpose AI being added on December 6, 2022).

199. *See, e.g.*, AI Act, *supra* note 2, recital 100 ("When a general-purpose AI model is integrated into or forms part of an AI system, this system should be considered to be general-purpose AI system when, due to this integration, this system has the capability to serve a variety of purposes. A general-purpose AI system can be used directly, or it may be integrated into other AI systems.").

200. *See id.* recital 97 ("The notion of general-purpose AI models should be clearly defined and set apart from the notion of AI systems to enable legal certainty . . . . Although AI models are essential components of AI systems, they do not constitute AI systems on their own. AI models require the addition of further components, such as for example a user interface, to become AI systems. AI models are typically integrated into and form part of AI systems. This Regulation provides specific rules for general-purpose AI models and for general-purpose AI models that pose systemic risks, which should apply also when these models are integrated or form part of an AI system.").

201. *Id.* art. 3(63). A "general-purpose AI system," means "an AI system which is based on a general-purpose AI model and which has the capability to serve a variety of purposes." *Id.* art. 3(66).

explicitly excludes "AI models that are used for research, development or prototyping activities before they are placed on the market."[202]

The Act separates general-purpose AI models into two distinct categories: those with and without "systemic risk." A general-purpose AI model has systemic risk if it either "has high impact capabilities evaluated on the basis of appropriate technical tools and methodologies, including indicators and benchmarks" or the Commission sees the model as equivalent to that.[203] "High impact capabilities," in turn, is a phrase defined in the law as "capabilities that match or exceed the capabilities recorded in the most advanced general-purpose AI models."[204] When the Commission decides whether a general-purpose model poses equivalent systemic risks, it must "take into account" criteria laid out in Annex XIII, which are largely technical criteria, including the size of the model, the amount of computation used to train the model, the different modalities of the model (text-to-text, image-to-text, etc.), and concerns about the model's reach—namely, the number of registered end-users.[205] The Act separately creates a presumption of systemic risk when over 10^{25} floating point operations (FLOPs) are used for training.[206] Annex XIII also creates a presumption where the model is available to ten thousand registered business users.[207] The Commission has the power to amend these thresholds over time.[208]

Providers of general-purpose AI models have certain obligations, with additional obligations if the AI model is deemed to have systemic risk. Providers must keep up-to-date technical documentation that can be reviewed by the AI Office and national authorities. They must also make information and documentation available to downstream providers of AI systems who intend to integrate the general-purpose AI model into their system.[209] These requirements are similar to the documentation requirements for high-risk

202. *Id.* art. 3(63).

203. *Id.* art. 51(1).

204. *Id.* art. 3(64).

205. *Id.* art. 51(1); *see also id.* Annex XIII. Article 52 contains more detail about the procedure of such a designation and appeals by the providers.

206. *Id.* art. 51(2). Notably, this FLOP-based threshold is similar to one in the Biden Executive Order on AI, except the number there is 10^{26} FLOPs, an order of magnitude higher. This means many more models meet the threshold in Europe than did in the US under Biden. Epoch.AI has studied documentation of more than 300 models, and has, as of May 12, 2025, found one model that would exceed the 10^{26} threshold, while about 20 exceed the 10^{25} line. EPOCH AI, DATA ON AI: AI MODELS (Aug. 12, 2025), https://epoch.ai/data/ai-models.

207. AI Act, *supra* note 2, Annex XIII.

208. *Id.* art. 51(3).

209. *Id.* art. 53(1)(b).

systems that must be made available for oversight, as well as the instructions to deployers that are needed to properly use the AI system.[210]

In addition, the Act requires providers of general-purpose AI models to do three things related to copyright law: (1) establish a copyright policy; (2) operationalize rightsholders' opt-outs from training datasets; and (3) draw up and make publicly available a sufficiently detailed summary about content used for training their model.[211] The Commission recently released a template for describing training data.[212] The Act exempts providers of open-source models from the requirements for general-purpose AI systems, but not from the copyright requirements.

Providers of general-purpose AI models with systemic risk must additionally perform model evaluation, conduct and document adversarial training, assess and mitigate possible systemic risks, report and document information about "serious incidents," and attend to cybersecurity protection.[213] The Act does not exempt open-source models with systemic risk.[214]

The Act's oversight framework for general-purpose AI models is distinct from that used for high-risk AI systems. (However, recall that a general-purpose model can end up used in a high-risk AI system, which we discuss below.[215]) For one, the conformity assessment framework—the central framework for the law overall—just doesn't apply to general-purpose AI models.[216] Instead the Act grants "exclusive powers to supervise and enforce" the relevant provisions to the AI Office of the Commission.[217] To support this,

210. *See supra* Section III.A.

211. AI Act, *supra* note 2, art. 53(1)(c) (citing Article 4(3) of Directive (EU) 2019/790, which is about the opt-out right).

212. *See* EUR. COMM'N, TEMPLATE FOR THE PUBLIC SUMMARY OF TRAINING CONTENT FOR GENERAL-PURPOSE AI MODELS (July 24, 2025), https://ec.europa.eu/newsroom/dae/redirection/document/118578; *Drawing-Up a General-Purpose AI Code of Practice*, EUR. COMM'N (Aug. 1, 2025), https://digital-strategy.ec.europa.eu/en/policies/ai-code-practice; EUROPEAN AI OFFICE, EUROPEAN AI OFFICE WORKING GROUP MEETINGS: CODE OF PRACTICE FOR GENERAL-PURPOSE AI: TEMPLATE FOR SUMMARY OF TRAINING DATA, EUR. COMM'N (Jan. 17, 2025), https://ec.europa.eu/newsroom/dae/redirection/document/111909; *Commission Presents Template for General-Purpose AI Model Providers to Summarise the Data Used to Train Their Model*, EUR. COMM'N: PRESS RELEASE (July 24, 2025), https://digital-strategy.ec.europa.eu/en/news/commission-presents-template-general-purpose-ai-model-providers-summarise-data-used-train-their.

213. AI Act, *supra* note 2, art. 55.

214. *Id.* art. 53(2).

215. *See also id.* recital 97.

216. Oddly, Articles 40 and 41 include statements that if general-purpose AI systems follow the standards or common specifications, respectively, they are presumed to be in conformity, despite the law nowhere *requiring* them to be in conformity. *Id.* arts. 40(1), 41(1).

217. *Id.* art. 88(1).

the Commission is granted the powers to request documentation and information, to evaluate models, to request remedial measures, and to make them binding after a "structured dialogue" with the provider.[218]

The Act also considers situations governing the overlap of general-purpose models and high-risk AI systems. Where any AI system is made by a provider of a general-purpose AI and incorporates it, the AI Office has the powers of a market surveillance authority.[219] Where a deployer substantially modifies a general-purpose AI system for use in a high-risk AI context (i.e., fine-tunes a generative AI), it is considered a provider of a high-risk system and falls under the Act's high-risk regulation discussed at length above.[220]

Where a national market surveillance authority has reason for concern that a general-purpose AI system can be used in an *unmodified* way by deployers, and that it violates the requirements of the Act, the authority is directed to cooperate with the AI Office to ensure compliance. If the national market surveillance authority is stonewalled by the general-purpose AI provider, it is directed to submit a request to the AI Office to enforce its right to necessary oversight information.[221]

C. AD HOC ELEMENTS AND OTHER STRANDS OF REGULATION

Finally, the Act contains ad hoc elements and other strands of regulation that fit neither into its core reliance on the NLF nor into its afterthought approach to general-purpose AI models. We conclude this Part by pointing to several of these, but again, this is not an exhaustive review. We discuss the Act's required disclosures to individuals; the Act's requirements of "Fundamental Rights Impact Assessments" and AI explanations to affected individuals; and strands of regulation regarding "innovation" that address regulatory sandboxes, real-world testing, and small businesses.

As mentioned above, the AI Act largely does not provide individual rights to affected persons, perhaps relying on rights established through other laws, such as the GDPR. However, Chapter IV of the Act (which consists of only one article, Article 50) establishes "Transparency obligations for providers and deployers of certain AI systems."[222] These include the following: AI systems, such as chatbots, must be designed to inform people that they are interacting

218. *Id.* arts. 91–93.

219. *Id.* art. 75(1).

220. *Id.* art. 25(1)(c). This is actually true for any substantial modification of any type of AI system—anyone who does it becomes a provider—but it is particularly likely to come up in the case of fine-tuning generative AI.

221. *Id.* arts. 75(2)–(3).

222. *Id.* art. 50.

with an AI system.[223] Deployers of an emotion recognition system must inform affected people that the system is in use.[224] So must deployers of biometric categorization systems.[225]

Several of Article 50's transparency provisions address synthetic content, or "deep fakes." The providers of general-purpose AI systems that generate synthetic content must make sure outputs are marked as AI-generated in a machine-readable way.[226] Deployers must disclose that images, audio, or video output has been artificially generated or manipulated.[227] They also must disclose if text on matters of public concern has been artificially generated or manipulated.[228]

Two other provisions of the AI Act appear to be ad hoc responses to critiques made during drafting by EU's data privacy regulators, the European Data Protection Board (EDPB) and European Data Protection Supervisor (EDPS).[229] These regulators criticized an earlier draft of the AI Act for failing to protect or really even address the rights of affected individuals.[230] Consequently, the EDPB and EDPS urged the drafters to establish rights and remedies for "individuals subject to AI systems."[231] Specifically, they called for a "right to explanation" of AI decisions,[232] which Article 86 of the AI Act consequently provides.[233] One of us has written detailed analysis of when and how this right applies, including how it interacts with similar rights established by the GDPR.[234]

Criticisms from the EDPB and EDPS also led to the adoption of risk mitigation at the deployer level: the Fundamental Rights Impact Assessment in Article 27. The EDPB and EDPS noted that initially, the AI Act (consistent

223. *Id.* art. 50(1) ("intended to interact directly with natural persons are designed and developed in such a way that the natural persons concerned are informed that they are interacting with an AI system").

224. *Id.* art. 50(3).

225. *Id.*

226. *Id.* art. 50(2).

227. *Id.* art. 50(4).

228. *Id.*

229. *See* EDPB-EDPS Joint Opinion, *supra* note 106.

230. *Id.* ¶ 18 ("Whether they are end-users, simply data subjects or other persons concerned by the AI system, the absence of any reference in the text to the individual affected by the AI system appears as a blind spot in the Proposal.").

231. *Id.* ("the EDPB and the EDPS urge the legislators to explicitly address in the Proposal the rights and remedies available to individuals subject to AI systems").

232. *Id.* ¶ 60 ("A right to explanation should provide for additional transparency.").

233. AI Act, *supra* note 2, art. 86.

234. Margot E. Kaminski & Gianclaudio Malgieri, *The Right to Explanation in the AI Act*, *in* THE EU ARTIFICIAL INTELLIGENCE ACT: A THEMATIC COMMENTARY (Gianclaudio Malgieri, Gloria González Fuster, Alessandro Mantelero & Gabriela Zanfir-Fortuna eds., forthcoming 2026).

with the NLF) placed risk mitigation requirements upon providers only, and not upon deployers.[235] They explained that sometimes AI system users (deployers) should be the ones conducting risk mitigation.[236] Consequently, the AI Act's drafters added Article 27. It requires deployers that are government entities, private entities providing public services, and deployers of high-risk AI systems that price insurance or determine creditworthiness, to conduct a Fundamental Rights Impact Assessment (FRIA) prior to first use.[237] The FRIA must among other things, identify the categories of persons and groups likely to be affected and the specific risks of harm to them, and outline measures taken in case risks materialize, including "internal governance and complaint mechanisms."[238] The results must be reported to the market surveillance authority.[239]

Finally, the AI Act contains numerous provisions that it describes as "measures in support of innovation."[240] We address these as separate strands of regulation, distinct from the NLF framework. They are not strictly speaking ad hoc, as they have been part of the Act's framework since early drafting. These strands include provisions establishing AI regulatory sandboxes, provisions on testing systems in real-world conditions, and provisions regarding the regulation of small businesses.

A regulatory sandbox is a type of temporary and experimental regulatory regime that might provide a break from regulation in exchange for supervision.[241] Typically, a sandbox entails public-private collaboration, coupled with clarifying guidance from a regulator. They were first used in the context of U.K. Fintech regulation.[242] As contemplated in the AI Act, regulatory sandboxes aim to foster innovation, support startups, improve legal certainty, contribute to the sharing of best practices, and contribute to "evidence-based regulatory learning."[243]

---

235. *See* EDPB-EDPS Joint Opinion, *supra* note 106, ¶¶ 20–21.

236. *Id.*

237. AI Act, *supra* note 2, art. 86(1)*; see also id.* Annexes III(5)(b)–(c) ("AI systems intended to be used to evaluate the creditworthiness of natural persons or establish their credit score, with the exception of AI systems used for the purpose of detecting financial fraud; AI systems intended to be used for risk assessment and pricing in relation to natural persons in the case of life and health insurance").

238. *Id.* arts. 86(1)(a)–(f).

239. *Id.* art. 86(3).

240. *See id.* ch. VI (titled "Measures in Support of Innovation"); *id.* art. 57 (titled "AI Regulatory Sandboxes").

241. *See generally* Sofia Ranchordás & Bart van Klink*, Special Issue Experimental Legislation in Times of Crisis*, 11 L. & METHOD 1 (2022).

242. *Id.* at 6.

243. AI Act, *supra* note 2, art. 57(9).

The AI Act requires that member states establish at least one AI regulatory sandbox by August 2026.[244] AI regulatory sandboxes are described as a "controlled environment . . . [that] facilitates the development, training, testing and validation of innovative AI systems for a limited time before their being placed on the market or put into service pursuant to a specific sandbox plan agreed between the providers . . . and the competent authority."[245] An AI provider that has been subject to a sandbox may use its sandbox "exit report" to later demonstrate compliance with the AI Act.[246]

An overlapping strand of regulation has to do with testing. Multiple provisions of the Act address testing AI systems in real-world conditions. Real-world testing is contemplated within regulatory sandboxes.[247] Article 60 additionally governs real-world testing of up to a year outside of the sandbox context, subject to a real-world testing plan that is submitted to the market surveillance authority.[248] Article 61 requires informed consent from test subjects.[249]

The AI Act also contains several measures that affect its application to small businesses and startups.[250] For example, it requires that small and medium-sized businesses be given priority access to regulatory sandboxes.[251] It also requires that member states facilitate participation by small businesses in the standardization process.[252] It requires that small businesses and startups be charged reduced fees for conformity assessments.[253] And the Act requires the Commission to develop guidelines for simplified compliance with the quality management system (required under Article 17) for businesses with fewer than ten employees generating under €2 million in revenue.[254]

IV. ANALYSIS

In this last Part, we offer some analysis. We begin with a discussion of whether the AI Act is better understood as an instantiation of the

244. *Id.* art. 57(1).
245. *Id.* art. 57(5).
246. *Id.* art. 57(7).
247. *Id.* art. 57(5).
248. *Id.* arts. 60(1), 60(4).
249. *Id.* art. 61(1) ("freely-given informed consent shall be obtained from the subjects of testing prior to their participation in such testing").
250. *Id.* art. 62.
251. *Id.* art. 62(1)(a).
252. *Id.* art. 62(1)(d).
253. *Id.* art. 62(2).
254. *See* Commission Recommendation, art. 2(3), 2003 O.J. (L 124) 39 (defining "microenterprise"), https://eur-lex.europa.eu/LexUriServ/ LexUriServ.do?uri=OJ:L:2003:124:0036:0041:EN:PDF.

precautionary principle, or a law promoting the uptake of AI. We then discuss how the Act constructs high-risk AI systems through a product-safety framing, and the consequences of doing so. We emphasize that the Act should be understood as the unique product of its drafting story. We describe it as a "legal exoskeleton": a hard-law framework with a softer-law interior that leaves many open questions. We close with some musings on global power politics and the future of the law and its global influence.

A. PRECAUTION, OR A BID FOR AI BUSINESS?

If there's one concept American lawyers tend to think of with respect to the difference between EU and U.S. regulators, it's the precautionary principle, which emphasizes caution about new technologies. In the United States, we innovate first, ask questions later, while in Europe the red tape is layered on thick—or so the story goes.[255]

For this reason, it is notable that the AI Act puts the "uptake of human-centric and trustworthy artificial intelligence" on an equal purposive footing to the "high level of protection of health, safety, fundamental rights" it aims to "ensure."[256] The Act centrally prioritizes the uptake of AI. One way to understand the AI Act, against the typical backdrop of EU regulation, is that the EU decided to attract more AI use.[257] Being the first in the world to pass a comprehensive AI law may have been a play to capture more of the global market for AI, with protection of EU citizens a secondary goal.[258]

This would certainly help explain the reliance on the product-safety framing and the shunting of fundamental rights concerns to a framework of self-certification. As one of us has argued elsewhere, there is a growing global convergence on risk regulation as governance of AI.[259] Risk regulation is

255. Douglas A. Kysar, *It Might Have Been: Risk, Precaution and Opportunity Costs*, CORN. L. FAC. PUBL'NS, PAPER 50, 1, 3–4 (2006) (noting United States favors cost-benefit analysis that predicts, weighs, and aggregates consequences of policy proposals to identify "welfare-maximizing uses of public resources," while EU approach to risk regulation is associated with precautionary principle). *But see* Jonathan B. Wiener, *Whose Precaution After All? A Comment on the Comparison and Evolution of Risk Regulatory Systems*, 13 DUKE J. COMPAR. & INT'L L. 207, 213–15 (2003) (noting the common perception that the EU favors precaution and the United States favors more permissive regulation is oversimplified).

256. AI Act, *supra* note 2, art. 1(1).

257. *See* Troels Krarup & Maja Horst, *European Artificial Intelligence Policy as Digital Single Market Making,* 10 BIG DATA & SOC'Y 1 (2023).

258. *See*, *e.g.*, Daniel Leufer, Fanny Hidvegi & Alessia Zornetta, *The Pitfalls of the European Union's Risk-Based Approach to Digital Rulemaking*, 71 UCLA L. REV. DISCOURSE 156, 166 (2024) ("[T]he Commission's strategy on AI since 2018 has focused primarily on boosting AI uptake across the EU, with measures to tackle the negative impacts of such uptake coming as a secondary consideration.").

259. Kaminski, *supra* note 51, at 1396.

typically "ex ante, systemic, and concerned with aggregate outcomes,"[260] rather than focused on ex post accountability for individualized harms. Among risk regulation's "policy baggage" is that risk regulation often "takes as its starting point that a technology must be fixed so that it can be used."[261] Thus, not only does the Act expressly state that it wants to increase uptake of AI, but by adopting a risk regulation framework, it essentially takes a stance that AI is fixable, but inevitable.

While risk regulation is not always incompatible with a precautionary approach—precaution and risk analysis often go hand-in-hand, after all—the legal and intellectual framework of this particular approach to risk regulation contrasts with the possibility of stronger prohibitions, a robust system of individual redress, and ex post liability for harms. While the AI Act does notably contain bans, they are narrow and limited. And the choice to create a law under the NLF framework is a break from what one might typically expect from an EU concerned with vindicating individual fundamental rights.

B. LEGALLY CONSTRUCTING HIGH-RISK AI SYSTEMS THROUGH PRODUCT SAFETY

The AI Act constructs the legal concept of "high-risk AI systems" through the values and institutions of product safety regulation.[262] It treats AI systems as though they are elevators or children's toys, rather than bureaucratic systems for decision-making.[263] Not only does the Act characterize AI systems as products, it envisions AI systems as though they are products designed for and used for a particular intended purpose. Inherently, general-purpose AI systems do not readily fit into this regulatory model.[264]

In characterizing AI systems as products, the Act imports the values of product safety regulation. It prioritizes protection of health and safety over the protection of fundamental rights.[265] It relies on ex post measures such as product recalls, which indicates that some level of harm is acceptable, up to a point. This is not how the EU generally treats fundamental rights. Typically,

260. *Id.* at 1369.

261. *Id.* at 1397.

262. For a discussion of the method of legal construction of technology, see Margot E. Kaminski & Meg Leta Jones, *Constructing AI Speech*, 133 YALE L.J. 1212 (2024). Here, we discuss the objects (AI as product), values (the values of product safety regulation), and institutions (the institutions of the NLF). *See also* Petit & Almada, *supra* note 57.

263. *See also* Veale & Zuiderveen Borgesius, *supra* note 57, at 102.

264. *See generally* Claire Boine & David Rolnick, *Why the AI Act Fails to Understand Generative AI*, 26 MINN. J. L., SCI. & TECH. 61 (2025).

265. Recall that AI systems with safety implications undergo third-party conformity assessments, while AI systems implicating fundamental rights get self-certification.

EU law turns on whether a right has been violated, not whether a certain level of measurable harm has been caused.[266]

The AI Act constructs AI systems as particular objects (products), through particular institutions (product safety regulators). The AI Act piggybacks on the existing institutional framework of the NLF (albeit as discussed with some important caveats). Its infrastructure relies on the existence of two classes of member state institutions: (a) market surveillance authorities, responsible for market surveillance and ensuring compliance, and (b) notifying authorities, responsible for designating third-party conformity assessment bodies. Recall that member states may designate existing NLF institutions for these roles. This matters because the Act will be primarily enforced by institutions trained in operationalizing product safety values rather than fundamental rights values.[267] For example, in Finland, the Transport and Communications Agency, a preexisting market surveillance authority, is responsible for the Act's enforcement.[268]

Member states do have discretion in deviating from these institutions, however. For example, Italy instead designated its national cybersecurity agency as its market surveillance authority.[269] Other member states have designated their data protection authority as the market surveillance authority.[270] Still others have created new bodies specifically for regulating AI systems.[271] The array of different institutions with different degrees and kinds of expertise, resources, and values will greatly affect how the Act is implemented in practice. In short, while some of these institutions will bring values other than products safety values to the enforcement table, there is no guarantee these institutions will be well-versed in the CJEU's fundamental

266. *See* Mireille Hildebrandt, *Beyond the GDPR?*, Presentation for the COHUBICOL ERC ADG Project 1, 18 (2023), https://www.cohubicol.com/assets/uploads/response-hildebrandt-purtova.pdf. *See also* Katerina Demetzou, Data Protection Impact Assessment (2014–24) (PhD dissertation, Radboud Business Law Institute) (discussing risks to rights under the GDPR).

267. *See* Veale & Zuiderveen Borgesius, *supra* note 57, at 112 ("The enforcement mechanism is a creature of product safety.").

268. *Overview of All AI Act National Implementation Plans*, EU ARTIFICIAL INTELLIGENCE ACT (Nov. 8, 2024), https://artificialintelligenceact.eu/national-implementation-plans/ ("A draft implementing act from October, 2024, appoints 10 already existing market surveillance authorities . . . . The Finnish Transport and Communications Agency will act as the single point of contact.").

269. *Id.* ("A legislative proposal from May 2024 designates the National Cybersecurity Agency (Agenzia per la Cybersicurezza Nazionale, ACN) as market surveillance authority with monitoring, inspection and enforcement powers in relation to AI systems").

270. *Id. See, e.g.*, Luxembourg & Malta.

271. Overview of All AI Act, *supra* note 268 (Poland and Romania sections).

rights decisions. This great variety of institutions will also lead to a great deal of heterogeneity in implementation across member states.

The AI Act does not stand in isolation, however, and it is important to be aware of what does and does not exist around it. For example, product safety regulation typically interacts with product liability, with liability providing a backstop to regulation and an opportunity for individual redress. However, there is no corresponding harmonized product liability law on the EU level. The proposed AI Liability Directive, which would have done some harmonizing on AI liability, is now dead.[272] Member states' approaches to liability accordingly will not be harmonized with respect to AI.

Without a backstop product liability directive, the Act's central reliance on risk regulation takes on more significance. That reliance on risk regulation affects both substance and institutional design. It results in a strange fit between the tools the Act uses and the harms it aims to prevent.[273] The AI Act turns on a quantified definition of "risk," where human rights violations are typically not quantifiable.[274] With only a few exceptions,[275] the Act does not afford specific individual rights. For a law purportedly aimed in no small part towards fundamental rights protection, this is bizarre.[276]

From an institutional design perspective, the lack of individual rights or redress is significant. It potentially makes it significantly harder to get cases on fundamental rights protections under the Act before the Court (the CJEU). A contrast with the GDPR here may be helpful. As outlined in Part I, typically, a case comes before the CJEU when it is referred by a member state's national court on a question of EU law. Under the GDPR, an individual whose rights are violated can lodge a complaint with a member state's data protection

272. *See* Caitlin Andrews, *European Commission Withdraws AI Liability Directive from Consideration*, INT'L ASS'N PRIV. PROS. (Feb. 12, 2025), https://iapp.org/news/a/european-commission-withdraws-ai-liability-directive-from-consideration (describing the withdrawal of the proposed directive in Feb. 2025); Cynthia Kroet, *Lawmakers Reject Commission Decision to Scrap Planned AI Liability Rules*, EURO NEWS (Feb. 18, 2025), https://www.euronews.com/next/2025/02/18/lawmakers-reject-commission-decision-to-scrap-planned-ai-liability-rules (showing attempts to revive the directive). *But see* Deimante Rimkute, *AI Liability After the AILD Withdrawal: Why EU Law Still Matters?*, OXFORD BUS. L. BLOG (Apr. 1, 2025), https://blogs.law.ox.ac.uk/oblb/blog-post/2025/04/ai-liability-after-aild-withdrawal-why-eu-law-still-matters (arguing that existing EU law on products liability nonetheless pushes towards harmonization).

273. Kaminski, *Regulating the Risks of AI*, *supra* note 51, at 1400.

274. *Id.* at 1401 ("A second, central problem of AI risk regulation is that the risks raised by AI systems are varied, not always quantifiable, often contested, and sometimes excruciatingly or even impossibly hard to define.").

275. *E.g.*, AI Act, *supra* note 2, art. 86, and some notification requirements, *id.* art. 50.

276. *See* EDPB-EDPS Joint Opinion, *supra* note 106, ¶ 18 ("Whether they are end-users, simply data subjects or other persons concerned by the AI system, the absence of any reference in the text to the individual affected by the AI system appears as a blind spot in the Proposal.").

authority, can sue a member state's data protection authority (DPA), or can sue a data controller or processor in court.[277] If that case goes up before the national court, that court can then refer questions on data protection law to the CJEU. (This process itself takes a long time!) This has in practice enabled the CJEU to act as the institutional rights-protective backstop to the regulatory regime outlined in the GDPR.[278]

The AI Act, by contrast, lacks as robust of a fundamental rights backstop. On the one hand, the Charter is still the source of fundamental rights and backstops the Act whether it is extensively operationalized or not.[279] On the other, the Act for the most part channels enforcement into market surveillance authorities and unlike the GDPR does not establish a right to sue those authorities, or to sue AI providers.[280] Whether or not it is possible to sue designated market surveillance authorities is a matter of member state law. This makes it potentially harder for the CJEU to serve as a rights-protective backstop to the AI Act's regime.[281]

277. *See* GDPR, *supra* note 19, arts. 77–79.

278. *See* Margot E. Kaminski & Meg Leta Jones, *American's Guide to the GDPR*, 98 DENV. L. REV. 93 (2021).

279. *See* Simona Demková, *The EU's Artificial Intelligence Laboratory and Fundamental Rights*, *in* REDRESSING FUNDAMENTAL RIGHTS VIOLATIONS BY THE EU 391, 411 (Melanie Fink ed., 2024) ("the AI Act will need to be applied in conjunction with the existing EU law, including the rules on remedies and existing data protection rules").

280. *Id.* at 409 ("it remains to be stressed that judicial remedies are rather limited in the context of AI-powered conduct based on composite administrative procedures involving actors at EU and Member State levels"); *id.* at 415–16 ("Section 4 in the final version of the Act . . . provides however only provides a limited consolidation of the calls for enhancing access to justice against the risks of AI . . . the remedies under the AI Act are essentially two-fold: (a) a product-related complaint mechanisms before the designated market surveillance authorities; (b) the right to an explanation of individual decision-making when the latter is made on the basis of a high-risk AI output").

281. Important caveat: with general-purpose AI, the Court will be able to hear questions about the Commission. *See Court of Justice of the European Union (CJEU): Overview*, EU, https://european-union.europa.eu/institutions-law-budget/institutions-and-bodies/search-all-eu-institutions-and-bodies/court-justice-european-union-cjeu_en ("ensuring the EU takes action . . . sanctioning EU institutions").

It is possible, however, that some AI Act questions will make it to the CJEU regardless,[282] including some provisions with direct effect,[283] or where individuals have rights to sue under other legal regimes.

Once again, however, the AI Act is not the only law governing AI in the EU. The Act depends on the existence of other laws that afford fundamental rights protections, consumer protections, and other requirements. These are the legal waters AI providers swim in. The Charter continues to provide protections for fundamental rights. The GDPR continues to afford individuals data protection rights. The Digital Services Act establishes specific obligations

282. Art. 86 may also come up as the one explicit individual remedy provided in the AI Act. *See, e.g.*, Request for a Preliminary Ruling from the Sofiyski Rayonen Sad (Bulgaria), 2025 O.J. (C-806/1080), https://eur-lex.europa.eu/eli/C/2025/1080/oj/eng (referring multiple questions about the AI Act to the CJEU in a case involving the Consumer Protection Directives) (for example: "Must Article 86(1) of Regulation (EU) 2024/1689 (1) be interpreted as meaning that the consumer has the right, within the meaning of Directives 2011/83/EU (2) and 93/13/EEC, (3) to know from the service provider how and with the aid of what elements [and] parameters automated decisions (invoices) were generated on the basis of data which the trader collected automatically in the context of a contract for the provision of mobile telecommunications services? . . . 4. [Must Art. 86] be interpreted as permitting the court to demand from the trader the black box data, the source code and the algorithm relating to the way in which automated decisions are made under the consumer contract? 5. Must Article 86(1) of Regulation (EU) 2024/1689, read in conjunction with Article 47 of the Charter of Fundamental Rights of the European Union, read in conjunction with Article 38 of the Charter, and with Directive 2011/83/EU, be interpreted as meaning that an automated decision generated by a trader under a contract with a consumer for mobile telecommunications services permits that automated decision to be reviewed by a human being, a judge, during real judicial proceedings? Must those provisions be interpreted as meaning that automated decisions . . . are subject to human review by a judge in real judicial proceedings? . . . 6. Must recitals 7 and 8 and Article 95(2)(a) of Regulation (EU) 2024/1689—the AI Act-and Directive 2011/83/EU . . . be interpreted as meaning that, where an automated decision-making system is operated and used . . . in the consumer contract, lawyers or senior judicial officers . . . with high moral and ethical standards must be involved in order to guarantee a transparent, effective and human-centric information system which takes account of fundamental rights? . . . 10. Must Article 5(1) of Directive 93/13/EEC and Article 86(1) of Regulation (EU) 2024/1689 be interpreted as meaning that the automatically generated invoices arising from a consumer contract . . . must be written in plain, intelligible language and the consumer has the right to demand an explanation from the trader as to how and by what algorithm the decision was made?").

283. *Direct Effect*, EUR. INDUS. RELS. DICTIONARY (Feb. 15, 2017), https://www.eurofound.europa.eu/en/european-industrial-relations-dictionary/direct-effect ("the CJEU identified three situations necessary to establish the direct effect of primary EU law. These are that: the provision must be sufficiently clear and precisely stated; it must be unconditional and not dependent on any other legal provision; it must confer a specific right upon which a citizen can base a claim."); *see also The Direct Effect of European Union Law*, EUR-LEX, https://eur-lex.europa.eu/EN/legal-content/summary/the-direct-effect-of-european-union-law.html ("in line with the general principles, this applies only under the condition that the rules are sufficiently clear, precise and relevant to the situation of the individual litigant (direct effect as clarified by the Politi v Ministero delle finanze Court judgement)").

for large online platforms. The Copyright in the Digital Single Market Directive addresses copyright concerns.[284]

There are two important consequences of the Act's reliance on other EU regulations as backdrop. First, the Act's reliance on existing fundamental rights protections in part justifies its product safety approach.[285] The Act purportedly can rely on risk regulation as its central mechanism precisely because individuals are afforded rights and redress through other regulations.

A second consequence of the Act's reliance on other EU law, however, is that if you have substantive questions about certain aspects of AI regulation, your answers may lie elsewhere. This certainly makes things trickier for American lawyers trying to assess the legality of a particular AI system. For example, if you want to determine the legality of a particular biometrics system, you will have to look both to the AI Act and to the GDPR.[286] If you want to determine how an online platform can use AI in content moderation, you will have to look to both the AI Act and the DSA. If you want to understand the EU's approach to copyright and training data, you will have to look to both the AI Act and EU copyright law.[287]

The relationship to data protection in particular is fraught. Both the AI Act and the GDPR are concerned with data, but towards fundamentally different ends. The GDPR at its core protects the "data subject"—individuals affected by the collection, processing, and use of personal data, protected not just under the GDPR but also under the Charter. The AI Act has no such concerns.[288] It's more concerned with mitigating inaccuracy and ensuring fit to purpose. However, the AI Act does not displace data protection law;[289] nor does it displace data protection institutions, on which it occasionally in fact relies.[290]

284. Directive (EU) 2019/790, of the European Parliament and of the Council (on copyright and related rights in the Digital Single Market), 2019 O.J. (L130) 92; *see also* Quintais, *supra* note 41.

285. *See* Eike Graef & Paul Nemitz, *Addressing the Challenge of Protecting Fundamental Rights Through AI Regulation in the European Union*, 71 UCLA L. REV. DISCOURSE 144, 151 (2024) ("It is important to look at the AI Act proposal together with these other elements, because the different initiatives and laws are designed to complement and strengthen each other.").

286. *See, e.g.*, Demková, *supra* note 279.

287. *See* Quintais, *supra* note 41.

288. *See* EDPB-EDPS Joint Opinion, *supra* note 106.

289. With the exceptions discussed above, in AI Act, *supra* note 2, arts. 10(5)(b), 59.

290. *See, e.g.*, *id.*, art. 70(9) (establishing the European Data Protection Supervisor as the market surveillance authority for EU providers).

C. THE ACT AS PRODUCT OF ITS DRAFTING STORY

The AI Act is the product of its messy drafting story. The Act may have started as a transplant of the NLF to high-risk AI systems, which itself is complex enough. But it ended up as a complicated Frankenstein.

The Act's approach to general-purpose AI was added after-the-fact. Several of its provisions on fundamental rights were added to respond to data protection regulators and other critics. And its bans on certain AI uses can almost all be traced to particular news items about particular applications of predictive AI. To understand the AI Act, then, you need to both be aware of its core reliance on the NLF, and be aware that as a law, it is quintessentially the product of its times.

It's also worth noting that while the AI Act took three years to draft, [291] many other EU regulations have had a far longer runway. Contrast, for example, the GDPR. European member states have had data protection law for decades, some since the 1970s. The EU-wide Data Protection Directive went into effect in the mid-1990s. The EU Charter, established in 2000, contains a fundamental right to data protection. By the time the GDPR was promulgated, there was both buy-in to and considerable infrastructure for the project of European data protection law. A similar point can be made about the slower progression in online platform regulation from the 2000 E-Commerce Directive to the 2022 Digital Services Act. By contrast, the AI Act was established as a regulation, over the course of only three years, with no preceding directive creating buy-in from citizens or member states. Contrasted with the EU's data protection regime and online platform regulation, it does not reflect a similar sort of bubble-up practical consensus developed over time. As a result, the AI Act may face legitimacy problems, leading to future amendments and/or difficulties with compliance.

D. THE ACT AS LEGAL EXOSKELETON

We describe the AI Act at its core as a *legal exoskeleton*: a hard-law legal framework surrounding a softer-law interior. The AI Act's central reliance on delegating—on having somebody else fill in most of the substantive blanks—means that at its core, it could turn out to be quite permissive. The Act delegates much of its substance to multiple potential actors, including technical standards-setting bodies and the Commission. Each bring with them democratic legitimacy problems, and a not insignificant likelihood of producing outcomes reflecting power politics. We here discuss both the delegation to technical standards-setting bodies, and the more recent

291. *Historic Timeline*, EU ARTIFICIAL INTELLIGENCE ACT, https://artificialintelligenceact.eu/developments/.

development of the Commission's General-Purpose AI (GPAI) Code of Practice.

A law can be softer or harder along different axes.[292] These include how mandatory versus voluntary a law is, who decides the rules, whether the rules are vague or specific, and more. The AI Act, like a number of recent European regulations, contains a complex and deliberate mixture of both harder and softer law. Its status as a regulation, rather than a directive, makes it directly binding on member states and thus harder law than a directive. Its enforcement mechanisms, which include large fines, also evidence a serious degree of hardness. However, the Act's central reliance on fuzzier language, instantiated in standards to be developed by technical standards-setting organizations, indicates a softer belly.

While the Act establishes a large and complicated accountability framework, we still don't know most of the actual substantive requirements for AI systems. The AI Act requires things like "the estimation and evaluation of the risks that may emerge"[293] and that "[t]raining, validation and testing data sets shall be relevant, sufficiently representative, and to the best extent possible, free of errors."[294] Both of these requirements, among many others, leave plenty of room for doubt about whether a particular provider is in compliance or not.

The main thing that providers of AI systems will want is certainty. Thus the rapid move to technical standards-setting. As soon as the joint technical committee of CEN/CENELEC finishes developing its AI Act standards, that document will probably become the de facto legal regime. Even though the standards that CEN/CENELEC come up with are not themselves technically binding law, they nonetheless effectively will have the force of law because compliance with them will establish a presumption of conformity with the law.

Until those standards are released, we do not know how substantive, detailed, or technical they will actually be. They may provide specific metrics or benchmarks, or they may echo the Act's existing fuzziness, or they may do both. At its core, then, the Act could end up with (a) less-than rigorous requirements negotiated in large part by private actors, (b) "technical" requirements that remain high-level and more like standards than like rules, or likely (c) some combination of the two.

292. *See* Kenneth W. Abbott & Duncan Snidal, *Hard and Soft Law in International Governance*, 54 INT'L ORG. 421, 421 (2000).

293. AI Act, *supra* note 2, art. 9(3).

294. *Id.* art. 10(3).

By delegating central decision-making, the Act at its core has democratic legitimacy problems.[295] One set of commentators has noted that by putting so central of an emphasis on technical standards-setting to govern fundamental rights, the AI Act puts a "constitutional bomb" under the NLF, a framework that has otherwise worked well to govern product safety in the EU.[296] As odd as it may be to rely on market surveillance authorities for oversight of fundamental rights when they lack such expertise, it is downright bizarre to delegate fundamental-rights decision-making to private standards-setting organizations.[297]

On the other hand, the EU's standards-setting organizations are not purely private, as they are in the United States. The EU's formal process for requesting technical standards and establishing them already involves more public sector decision-making and oversight than U.S. incorporation of private standards.[298] The CJEU has case law establishing that standards that are given legal effect are to be treated more like actual law.[299] For example, the CJEU has required that technical standards under copyright law be publicly accessible because they have the force of law.[300] Moreover, the AI Act imposes certain participation requirements on the AI standards-setting process in particular.[301]

---

295. *See* Marta Cantero Gamito & Christopher T. Marsden, *Artificial Intelligence Co-Regulation? The Role of Standards in the EU AI Act*, 32 INT'L J. L. & INFO. TECH. 1 (2024).

296. Veale & Zuiderveen Borgesius, *supra* note 57, at 105.

297. One commentator, speaking of parallel processes at NIST in the United States, has called these kinds of AI ethical governance documents crafted through technical standards-setting organizations "un-standards." Bryan H. Choi, *NIST's Software Un-Standards*, 9 GEO. L. TECH. REV. 65 (2025).

298. *See* Emily S. Bremer, *American and European Perspectives on Private Standards in Public Law*, 91 TULANE L. REV. 325 (2016).

299. Case C-588/21, Public.Resource.Org, Inc. and Right to Know CLG v Eur. Comm'n, ECLI:EU:C:2024:201, ¶ 80 (Mar. 2024) ("In the light of the foregoing considerations, it must be held, in accordance with the case-law referred to in paragraph 70 of the present judgment, that the requested harmonised standards form part of EU law.").

300. *Id.* ¶ 85 ("In those circumstances, it must be held that there is an overriding public interest, within the meaning of the last clause of Article 4(2) of Regulation No 1049/2001, justifying the disclosure of the requested harmonised standards.").

301. *See, e.g.*, AI Act, *supra* note 2, art. 40(3) ("The participants in the standardisation process shall seek to promote investment and innovation in AI, including through increasing legal certainty, as well as the competitiveness and growth of the Union market, to contribute to strengthening global cooperation on standardisation and taking into account existing international standards in the field of AI that are consistent with Union values, fundamental rights and interests, and to enhance multi-stakeholder governance ensuring a balanced representation of interests and the effective participation of all relevant stakeholders in accordance with Articles 5, 6, and 7 of Regulation (EU) No 1025/2012."); *id.* art. 62(d); *id.* recital 121 ("A balanced representation of interests involving all relevant stakeholders in the development of standards, in particular SMEs, consumer organisations and environmental and social stakeholders in accordance with Articles 5 and 6 of Regulation (EU) No 1025/2012 should therefore be encouraged"); *id.* recital 143.

The Act contemplates that the Commission can reject standards that it finds do not comport with the law and instead craft its own "common specifications."[302] That is, it backstops softer law with a public law option. (However, Commission-made law may itself be problematic—more on this in a moment.)

Understanding the AI Act as legal exoskeleton brings us to a crucial observation: the Act's central reliance on technical standards makes it vulnerable to international realpolitik. On the one hand, the EU passed the AI Act ostensibly to develop AI with built-in European values. On the other, by centrally relying on technical standards, the Act opens a side door to international influence. Other standards-setting organizations have already promulgated AI standards, or are far along in the process. China has recently increased its participation in international standards-setting.[303] In the United States, NIST issued its AI risk management framework in 2023,[304] and has turned its attention specifically to influence global standards-setting.[305]

Whatever standards the European Standardization Organizations issue will have an eye to international consensus-building. (In fact, they have to, under the WTO, to the extent anybody still follows the rules of international trade.[306])

---

302. AI Act, *supra* note 2, art. 41(1)(a)(iii) (when "(iii) the relevant harmonised standards insufficiently address fundamental rights concerns"); *see also id.* art. 40(2).

303. *See generally* Marta Cantero Gamito, *The Influence of China in AI Governance Through Standardisation*, 47 TELECOMM. POL'Y 102673 (2023).

304. *AI Risk Management Framework*, NAT'L INST. OF STANDARDS & TECH., https://www.nist.gov/itl/ai-risk-management-framework.

305. Jesse Dunietz, Elham Tabassi, Mark Latonero & Kamie Roberts, *A Plan for Global Engagement on AI Standards*, NAT'L INST. OF STANDARDS & TECH. (July 26, 2024), https://www.nist.gov/publications/plan-global-engagement-ai-standards; *see also Winning the Race: America's AI Action Plan*, WHITE HOUSE (July 2025), https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf (Pillar III: Lead in International AI Diplomacy and Security, including "Counter Chinese Influence in International Governance Bodies": "leverage the U.S. position in international diplomatic and standard-setting bodies to vigorously advocate for international AI governance approaches that promote innovation, reflect American values, and counter authoritarian influence").

306. *See* World Trade Organization, Agreement on Technical Barriers to Trade (TBT), ¶ 2.2, Apr. 15, 1995, https://www.wto.org/english/docs_e/legal_e/17-tbt.pdf ("Members shall ensure that technical regulations are not prepared, adopted or applied with a view to or with the effect of creating unnecessary obstacles to international trade."); *see also id.* ¶ 2.4 ("Where technical regulations are required and relevant international standards exist or their completion is imminent, Members shall use them, or the relevant parts of them, as a basis for their technical regulations except when such international standards or relevant parts would be an ineffective or inappropriate means for the fulfilment of the legitimate objectives pursued . . . ."). *See generally Technical Information on Technical Barriers to Trade*, WORLD TRADE ORG., https://www.wto.org/English/tratop_e/tbt_e/tbt_info_e.htm.

The AI Act itself notes this dynamic.[307] There are signs that the Commission is aware of this risk themselves; they purportedly left out one of the European Standardization Organizations out of fear about U.S. and Chinese influences on the process.[308]

There are first-mover advantages in the standards-setting race. The EU may have been the first to pass omnibus AI hard law, but it was not the first to promulgate standards.[309] This means that it is as likely to import U.S. and Chinese values via standards-setting as it is to effectively establish European AI standards for export. Again, the Commission's ability to reject and thus check delivered standards may be crucial in determining which way the influence flows. However, unlike the Court, the Commission is not a human rights body. The Commission is less incentivized than, e.g., the Court, to push back on inadequate standards for fundamental rights reasons.[310]

We end with a related recent plot twist concerning the recently developed GPAI Code of Practice. Again, there are many open questions about how this will all play out in the longer run. But the recent focus on the Code of Practice suggests a sort of shell game in where the Act's soft-law lawmaking may be taking place—and that the Commission is also not immune from political pressures.

Article 56 of the Act establishes a policy-making-qua-convening role for the AI Office at the Commission, with respect to general-purpose AI

307. *See* AI Act, *supra* note 2, art. 40(3) ("to contribute to strengthening global cooperation on standardisation and taking into account existing international standards in the field of AI that are consistent with Union values, fundamental rights and interests").

308. Luca Bertuzzi, *Commission Leaves European Standardisation Body Out of AI Standard-Setting*, EURACTIV (Dec. 7, 2022), https://www.euractiv.com/section/tech/news/commission-leaves-european-standardisation-body-out-of-ai-standard-setting/ ("[T]he European Commission set out its strategy to become more assertive in the way it participated in standard-setting, where it considered that non-European companies, particularly American and Chinese, have gained the upper hand. The strategy came as a slap in the face to ETSI, which the Commission accused of being held hostage by non-European influences and requested internal reform to give more weight to national standardisation bodies."); *see also* AI Act, *supra* note 2, art. 40(3) ("taking into account existing international standards in the field of AI that are consistent with Union values, fundamental rights and interests").

309. Technically, the Colorado AI Act passed first, but it is primarily an antidiscrimination law, not as omnibus.

310. *See, e.g.*, the saga of *Schrems I* and *Schrems II*, in which the Court twice found that the Commission's negotiated Safe Harbor agreement with the U.S. government violated fundamental rights under the Charter. Case C-362/14, Maximillian Schrems v. Data Protection Commissioner, ECLI:EU:C:2015:650 (Oct. 6, 2015) [*Schrems I*]; Case C-311/18, Data Protection Commissioner v. Facebook Ireland Ltd., ECLI:EU:C:2020:559 (July 16, 2020) [*Schrems II*].

models.[311] It tasks the AI Office with "encourag[ing] and facilitat[ing] the drawing up of codes of practice at Union level in order to contribute to the proper application of this Regulation, taking into account international approaches."[312] Consequently, the AI Office convened a group of experts to come up with the draft GPAI Code of Practice,[313] which was subsequently ratified by the Commission.[314]

Compliance with the GPAI Code of Practice serves to demonstrate compliance with the Act, at least until the standards-setting organizations arrive at a harmonized standard for general-purpose AI.[315] In fact, it appears that a general-purpose AI model provider could choose its own legal adventure and comply with the GPAI Code of Practice *instead of* any harmonized standard.[316] So if the GPAI Code of Practice is comparatively weak, while the technical-standards-setting output is more rigorous, the Code could provide a path of least regulatory resistance.

Unsurprisingly, the availability of this option led to "intense lobbying."[317] The Code of Practice was drafted involving over one thousand stakeholders from academia, civil society, and industry. But GPAI companies had a "special seat at the table," including privileged access to the latest version of the text.[318] Nonetheless, multiple U.S. companies indicated they would not sign on to the Code.[319] Ultimately, the Commission withstood the pressure, and both the

311. AI Act, *supra* note 2, art. 56(2) ("at least the obligations provided for in Articles 53 and 55").

312. *Id.* art. 56(1).

313. *The General-Purpose AI Code of Practice*, EUR. COMM'N (Sep. 9, 2025), https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai (characterizing the GPAI Code of Practice as a "voluntary tool, prepared by independent experts in a multi-stakeholder process, designed to help industry comply with the AI Act's obligations for providers of general-purpose AI models.").

314. *Commission Opinion on the Assessment of the General-Purpose AI Code of Practice*, EUR. COMM'N (Aug. 1, 2025), https://digital-strategy.ec.europa.eu/en/library/commission-opinion-assessment-general-purpose-ai-code-practice.

315. AI Act, *supra* note 2, art. 55(2) ("Providers of general-purpose AI models with systemic risk may rely on codes of practice within the meaning of Article 56 to demonstrate compliance with the obligations set out in paragraph 1 of this Article, until a harmonised standard is published.").

316. *Id.* ("Providers of general-purpose AI models with systemic risks who do not adhere to an approved code of practice or do not comply with a European harmonised standard shall demonstrate alternative adequate means of compliance for assessment by the Commission.").

317. Paul Nemitz & Amin Oueslati, *How US Firms Are Weakening the EU AI Code of Practice*, TECH POL. PRESS (June 30, 2025), https://www.techpolicy.press/how-us-firms-are-weakening-the-eu-ai-code-of-practice/.

318. *Id.*

319. Gian Volpicelli & Kurt Wagner, *Meta's Kaplan Signals Pushback Against EU Regulation for AI*, BLOOMBERG (Feb. 4, 2025), https://www.bloomberg.com/news/articles/2025-02-04/meta-s-kaplan-signals-pushback-against-eu-regulation-for-ai.

Commission and the AI Board finalized the process that put the Code in place.[320] But this story shows that companies try to find where the path of least resistance is, institutionally.

E. IN WHICH IT ALL COMES DOWN TO POWER POLITICS

This tracing of power politics brings us to our final point. What happens if U.S. and Chinese AI companies decide to ignore the Act?[321] Arguably, there are requirements in the Act with which it is impossible for providers of generative AI to comply. Will companies forego the EU market, or will EU regulators water down or fail to enforce the law?

Another way of framing this question is to ask what role European regulators will play in the development and deployment of technology this time around. European data protection law famously has been exported around the world.[322] But the AI Act is different. First, the AI Act enters the world stage at a very different standing than European data protection law. EU data protection law, as discussed, had substantial historic legitimacy within member states. Moreover, it involved repeat existing players, including regulated companies accustomed to its values, institutions, and mechanisms. Second, data protection law was deliberately designed for export: it contains the (in)famous "adequacy" mechanism, in which Europe will not export EU persons' data unless a country has been found to have adopted adequate data protection law.[323] (The United States represents a notorious exception.)

The AI Act, by contrast, was imposed top-down. There is increasing internal EU pressure to deregulate, to be more competitive.[324] The AI Act contains no adequacy mechanism; it relies, instead, on the harmonizing effects of technical standards. Those come, as mentioned, with their own map of power politics, including initial salvos by the United States and increased involvement by China. And the Trump administration, including the erstwhile ally Elon Musk[325] and AI booster JD Vance, have been playing transnational

320. *Commission Opinion*, *supra* note 314.

321. Professor Selbst calls this "the Principle of 'Fuck You.'"

322. *See* ANU BRADFORD, THE BRUSSELS EFFECT (2020); Anupam Chander, Margot E. Kaminski & William McGeveran, *Catalyzing Privacy Law*, 105 MINN. L. REV. 1733 (2021).

323. Paul M. Schwartz, *Global Data Privacy: The EU Way*, 94 N.Y.U. L. REV. 771 (2019); *see also* GDPR, *supra* note 19, art. 45 (discussing transfer to third countries).

324. MARIO DRAGHI, REPORT ON THE FUTURE OF EUROPEAN COMPETITIVENESS (2024) (commonly referred to as the "Draghi report"), https://commission.europa.eu/topics/eu-competitiveness/draghi-report_en.

325. Adam Satariano, *E.U. Prepares Major Penalties Against Elon Musk's X*, N.Y. TIMES (Apr. 3, 2025), https://www.nytimes.com/2025/04/03/technology/eu-penalties-x-elon-musk.html.

realpolitik with European regulators.[326] This time around, there may not be a Brussels Effect in which Europe exports its values.[327] Rather, the AI Act risks being gutted through standards from the inside out—or, being gutted in implementation or amended.

V. CONCLUSION

The AI Act already feels like a regulation from another era. It is long, complex, and grounded in a legal regime unfamiliar to a U.S. audience: the NLF. By framing AI systems through product safety law, EU lawmakers aimed to encourage the uptake of AI in the EU. But they ended up creating an instrument—what we've called a legal exoskeleton—that could end up hollowed out through its center, leaving real human rights work to other EU laws and institutions. Time will tell.

326. J.D. Vance, *Remarks by the Vice President at the Artificial Intelligence Action Summit in Paris, France*, AM. PRESIDENCY PROJECT (Feb. 11, 2025), https://www.presidency.ucsb.edu/documents/remarks-the-vice-president-the-artificial-intelligence-action-summit-paris-france ("[W]e believe that excessive regulation of the AI sector could kill a transformative industry just as it's taking off, and we'll make every effort to encourage pro-growth AI policies . . . . [W]e need international regulatory regimes that foster[] the creation of AI technology, rather than strangle[] it. And we need our European friends, in particular, to look to this new frontier with optimism rather than trepidation." (cleaned up)).

327. *See* Marco Almada & Anca Radu, *The Brussels Side-Effect: How the AI Act Can Reduce the Global Reach of EU Policy*, 25 GER. L.J. 646 (2024).

RECENTERING PUBLIC VALUES IN AI GOVERNANCE: EXAMPLES FROM THE BIDEN ADMINISTRATION

Deirdre K. Mulligan[†] *& Kenneth A. Bamberger*[††]

ABSTRACT

This Article situates key Biden-Harris Administration AI initiatives within a "governance-by-design" framework—an approach we previously developed that centers public values, sectoral expertise, and participatory policymaking in decisions to regulate through technology. Governance-by-design argues for reorienting AI governance around three core principles: (1) privileging human and public rights by empowering domain-specific agencies while building a shared set of tools and approaches for risk assessment; (2) expanding agencies' technical expertise through public hiring and multisector collaboration; and (3) preserving the publicness of policymaking through designs that foreground embedded values and embedding stakeholder engagement and impact evaluation throughout AI system development and deployment.

The Article uses three examples to illustrate how key Biden-Harris Administration AI actions reflect these governance-by-design principles:

- the Administration's layered regulatory strategy that empowers sectoral agencies to safeguard human rights and public safety in AI use;
- the expansion of AI and rights-based expertise within government and the establishment of collaborative structures for risk management and evaluation; and,
- the institutionalization of practices that surface and interrogate the normative assumptions embedded in AI systems, while scaffolding public participation throughout their lifecycle.

We argue that together these initiatives offer an alternative to prevailing AI governance debates—particularly the dichotomy between risk-based and rights-based approaches, and the call for a centralized AI regulator. Instead, such governance-by-design provides a field-centric model that leverages existing institutional capacities, protects democratic norms, and re-centers the public in the often-private domain of AI development. It offers a durable, epistemically responsible framework for regulating AI systems in a way that supports both human rights and legitimate democratic governance.

TABLE OF CONTENTS

DOI: https://doi.org/10.15779/Z38416T25K

† Professor, UC Berkeley School of Information; Co-Faculty Director, Berkeley Center for Law & Technology; former Principal Deputy U.S. Chief Technology Officer at the White House Office of Science and Technology Policy, and Director of the National Artificial Intelligence Initiative Office (NAIIO), 2023–2024.

†† The Rosalinde and Arthur Gilbert Foundation Professor of Law, UC Berkeley; Co-Faculty Director, Berkeley Center for Law & Technology.

Much gratitude to Rachel K. Mucha for her superb research assistance.

I. INTRODUCTION

Artificial Intelligence (AI) design and deployment displays attributes of a type that persistently confounds public governance. AI is not amenable to traditional command-and-control regulation reliant on uniform ex ante rules requiring certain conduct. Technical expertise resides largely in private rather than public actors and institutions. Engineers in the private sector make granular decisions regarding AI design. Private firms often manage the deployment of AI even when used by governments. Model developers and deployers, as well as workers and the public who use or whose rights and interests are affected by AI, possess the information needed to understand the

ways in which AI affects different segments of our society. And the nature of AI itself compounds the challenges of opacity, comprehension, and uncertainty that threaten to turn "over key policy questions to privately developed algorithmic systems."[1]

How, then, can we build a regulatory system that doesn't outsource policy decisions? One that is equipped to meaningfully protect rights and safety. One that withstands the corrosive and insidious way private sector processes can dilute or undermine public goals to fit within management practices. A regulatory system in which technological opacity doesn't prevent agency experts from applying their knowledge or preclude the public from participation.

In sum, how can we recenter public values in AI governance?

Elsewhere, we have taken a hard look at efforts to enlist technology to protect values in information and communication technology.[2] We concluded that "governance-by-design"—the purposeful effort to use technology to embed values or policies—had become a central mode of policymaking, but also that our existing regulatory system was fundamentally ill-equipped to prevent that phenomenon from subverting public governance.[3]

Specifically, we provided examples that showed how governance-by-design had undermined important governance norms and chipped away at our voting, speech, privacy, and equality rights.[4] We further described the structural limitations of traditional legal and governance bodies that contributed to this problem. These include the limited technical expertise of many policy making bodies; the absence of a venue for policymakers to have the meta-discussion about when and whether it is appropriate to enlist technology in the service of values at all; and, relatedly, if technology is to be so used, how to prioritize among values.[5]

These structural limitations, we explained, contributed to processes that subvert fundamental democratic norms of intentional, deliberative,

1. Deirdre K. Mulligan & Kenneth A. Bamberger, *Procurement as Policy: Administrative Process for Machine Learning*, 34 BERKELEY TECH. L.J. 781, 807 (2019) [hereinafter Mulligan & Bamberger, *Procurement as Policy*].

2. *See generally* Deirdre K. Mulligan & Kenneth A. Bamberger, *Saving Governance-By-Design*, 106 CALIF. L. REV. 697 (2018) [hereinafter Mulligan & Bamberger, *Governance-By-Design*]; Mulligan & Bamberger, *Procurement as Policy*, *supra* note 1.

3. Mulligan & Bamberger, *Governance-By-Design*, *supra* note 2, at 697; *see also* Kenneth A. Bamberger, *Technologies of Compliance: Risk and Regulation in a Digital Age*, 88 TEX. L. REV. 669, 675–76 (2010) (describing the ways that "the use of technology systems to hardwire compliance" can "raise what might be called administrative-law concerns—concerns regarding the subversion of public norms requiring transparency, public oversight, and accountability in the exercise of regulatory discretion.").

4. Mulligan & Bamberger, *Governance-By-Design*, *supra* note 2, at 698.

5. *Id.* at 750.

participatory, and expert public decision-making, free from capture or caprice. And they produced overbroad "technological fixes" that privilege singular values and often disfavor human rights.[6] We further argued that the use of technology to regulate without addressing the limitations of existing legal policy and enforcement processes. This allowed governments and private actors to mask their aims, led to both unwitting and intentional privileging of some values over others, and obscured the resulting policy outcomes from ongoing public scrutiny as they recede into the technical desiderata of the technical environment.[7]

Finally, we proposed a framework for "saving" governance-by-design that emphasized a set of approaches that together could align technology governance with the norms of public governance and therefore be more responsive to public values, and not just private interests.[8]

Three are regulatory principles:

(1) Privileging Human and Public Rights;

(2) Ensuring that Regulators Possess the Right Attributes, including Broad Authority and Competence, as well as technical expertise; and

(3) Maintaining the Publicness of Policymaking.[9]

The fourth is a design principle in service of the others, counseling modesty and restraint in design that wherever possible preserves flexibility rather than fixing values.[10]

This Article places that framework within the context of current debates regarding the appropriate metrics and institutional structure for AI governance: specifically (1) disputes over whether AI regulation should be "risk-based" or "rights-based," and (2) arguments that a new, dedicated regulatory body should be created to administer the governance of AI. It identifies the ways those arguments are misdirected and explains the ways that our technology and design governance principles offer an alternative emphasis. With these three regulatory principles in mind, this Article considers and frames examples of AI governance approaches taken by the Biden-Harris Administration. Those examples follow regulatory principles that seek to ensure AI is developed and implemented in ways that support both democratic values and human rights, as well as the public's safety and security: that the "public" is recentered in the often-private endeavor of AI development and

6. *Id.* at 739.
7. *Id.* at 721.
8. *Id.* at 705.
9. *Id.*
10. *Id.*

implementation. They reflect a field-centric, in contrast to what one of us has called "model-centric," model of AI governance. This approach centers and maintains the meaning-making processes, rules, and norms that guide interactions and decisions within fields and protects them against the epistemological and other displacements that too often are a byproduct of automating or informating[11] activities. They thereby "support epistemically responsible behaviour."[12] The approach addresses the inherent shortcomings of government technical expertise while at the same time reasserting agency domain expertise, and the rights and logics that undergird legitimate public governance processes.

These Biden-Harris Administration initiatives, then, concretize a suite of regulatory approaches for successfully recentering public values in AI governance. Such public values include (1) privileging human and public rights by maintaining expert agency authority while building a shared knowledge base for risk assessment methods and practices—a method for appreciating risk, while identifying which rights we seek to privilege; (2) bringing expertise and capacity into government through direct hiring and stakeholder involvement; and (3) maintaining the publicness of policymaking by evaluating impacts not technical systems and requiring stakeholder participation throughout the process.

II. THE MISDIRECTION OF AI GOVERNANCE DEBATES

Much of the current discourse around AI governance involves two debates. The first involves a binary discourse over the proper mode of AI regulation: specifically, whether AI should either be regulated in a manner that is focused on risks or, by contrast, one that places attention on rights. The second debate engages the question of governance institutions: whether addressing the risks of AI would be best accomplished through a new agency with technical expertise dedicated to regulating AI models or systems.[13]

---

11. SHOSHANA ZUBOFF, IN THE AGE OF THE SMART MACHINE: THE FUTURE OF WORK AND POWER (Basic Books 1988).

12. Judith Simon, *Distributed Epistemic Responsibility in a Hyperconnected Era*, *in* THE ONLIFE MANIFESTO: BEING HUMAN IN A HYPERCONNECTED ERA 145, 155–58 (Luciano Floridi ed., 2015).

13. An examination of the various proposals to create new U.S. agencies to regulate digital markets is outside the scope of this paper. For examples of such proposals, see Harold Feld, *The Case for the Digital Platform Act: Market Structure and Regulation of Digital Platforms*, ROOSEVELT INST. 17 (May 2019) (urging the United States to "either empower an existing agency or create a new agency to use these powers as necessary"); Tom Wheeler, Phil Verveer & Gene Kimmelman, *New Digital Realities; New Oversight Solutions in the U.S.: The Case for a Digital Platform Agency and a New Approach to Regulatory Oversight*, SHORENSTEIN CTR. 2 (Aug. 2020) (advocating for the creation of "a new Digital Platform Agency" to take an "agile approach to oversight built on risk management" that would include "cooperatively developed and

A. THE RISKS VS. RIGHTS BINARY: HOW TO GOVERN

Margot Kaminski has documented a "growing convergence" around the use of risk-based frameworks for AI governance.[14] Regulators, policymakers, and scholars have identified this suite of approaches as particularly appropriate in regulatory contexts in which the implementation of policy goals is "technically and legally opaque."[15] Requiring assessments of systemic risk level by context promises rigor in allocating regulatory focus,[16] as at least "[i]n its idealized form, risk-based regulation offers an evidence-based means of targeting the use of resources and of prioritizing attention to the highest risks in accordance with a transparent, systematic, and defensible framework."[17] Discerning levels of risk by context, moreover, allows regulators to choose to take on different levels of risk in light of broader benefits to society.[18] The EU AI Act's call for a "risk-based approach" in order to effect a "proportionate" set of binding rules for AI systems, for example,[19] reflects its "underlying objective" "to strike an optimal (or proportionate) balance between innovation and the benefits of AI systems on the one hand, and the protection of fundamental values such as safety, health and fundamental rights on the

enforceable code of conduct for specific digital activities"), https://shorensteincenter.org/wp-content/uploads/2020/08/New-Digital-Realities_August-2020.pdf; Digital Platform Commission Act of 2023, S. 1671, 118th Cong. (2023) (creating an expert federal agency to comprehensively regulate digital platforms to protect consumers, promote competition, and defend the public interest). For an overview of calls for the creation of new agencies to regulate privacy and a wide range of other platform and AI-related issues (search, robotics, algorithms, etc.), see Asad Ramzanali, Toward a Privacy Agency: Policy and Politics Appendix V (Apr. 6, 2021) (M.P.P. Policy Analysis, Harvard Kennedy School of Government), https://www.dropbox.com/scl/fi/zfw6bavrlxsd2r3vnf0gi/Toward-a-Privacy-Agency-Policy-and-Politics-PAE-Asad-Ramzanali-4.6.21.pdf?rlkey=ihibb9qpkcd0b5u8g7cfl8ty9&e=1&st=45cq8jmt&dl=0.

14. Margot E. Kaminski, *Regulating the Risks of AI*, 103 B.U. L. REV. 1347, 1347 (2023).

15. *Id.* at 1365–66.

16. *See EU AI Act: First Regulation on Artificial Intelligence*, EUROPEAN PARLIAMENT, https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence#:~:text=AI%20regulation%20in%20Europe%3A%20the%20first%20comprehensive%20framework,-In%20April%202021&text=AI%20systems%20that%20can%20be,or%20less%20AI%20compliance%20requirements (last updated Feb. 19, 2025) ("In April 2021, the European Commission proposed the first EU artificial intelligence law, establishing a risk-based AI classification system. AI systems that can be used in different applications are analysed and classified according to the risk they pose to users. The different risk levels mean more or less AI compliance requirements.").

17. Julia Black & Robert Baldwin, *Really Responsive Risk-Based Regulation*, 32 L. & POL'Y 181, 181 (2010).

18. *See generally* Kaminski, *Regulating the Risks of AI*, *supra* note 14.

19. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 12 July 2024 on artificial intelligence, recital 26, O.J. (L 1689).

other."[20] In a regulatory arena in which the prescription of detailed mandates is unfeasible, moreover, risk-based regulation's focus on outcomes and assessments leaves flexibility in how goals are met. This approach thus often allows organizations to tailor their compliance measures to specific risks and contexts, enlisting them to document their process decisions and success or failure at doing so.

By contrast, rights-based approaches foreground concerns related to fundamental rights and individual fairness. Rights-based AI governance frameworks treat rights such as nondiscrimination and privacy as fundamental and inviolable.[21] Legal scholars and policymakers in the rights-based camp point favorably to regulatory frameworks like the EU's GDPR, which follows a "binary logic." They provide a "minimum and non-negotiable level of protection" against certain harms of AI for all individuals, and a regulated entity's action either provides this protection or fails to do so.[22] Rights-based frameworks apply the same rules to everyone irrespective of the level of risk or harm. In the context of the GDPR, for example, a data processing action either provides users adequate protection from risk or harm or it falls short—there is no balancing of interests or level of harm that is tolerable at the individual or systemic level.[23]

We and other scholars of regulation have cautioned against the excesses of both binary approaches to technology governance. While risk-based regulation is valuable in contexts featuring easily measured harms, reliance on a "comprehensive, defined, ex ante, body of regulatory mandates"[24] to govern technology design can leave unanswered the question of *risk to what values* in contexts involving "big, often-unquantifiable, often-contested, often-contextual, and often-individualized 'risks.'"[25] As law-and-technology scholars Julie Cohen and Ari Waldman explain, regulatory oversight premised on an ex ante focus on risk mitigation can lead to a form of "regulatory managerialism" that embraces a narrow toolkit of risk modeling, digital control systems, and data analytics that ultimately focuses narrowly on efficiency and process values.[26] Regulating private activity around technology with these operational

20. Martin Ebers, *Truly Risk-Based Regulation of Artificial Intelligence: How to Implement the EU's AI Act*, 16 EUR. J. RISK REGUL. 684, 685 (2025).

21. *Fundamentals of a Human Rights-Based Approach to Generative AI*, BSR (Feb. 2025), https://www.bsr.org/files/BSR-Fundamentals-of-a-Human-Rights-Based-Approach-to-Generative-AI.pdf.

22. RAPHAËL GELLERT, THE RISK-BASED APPROACH TO DATA PROTECTION 2 (2020).

23. *Id.* An alternate take on the GDPR is that it is focused on a set of actions that were a priori determined to be high-risk.

24. Mulligan & Bamberger, *Governance-By-Design*, *supra* note 2, at 739.

25. Kaminski, *Regulating the Risks of AI*, *supra* note 14, at 1378–79.

26. Julie E. Cohen & Ari Ezra Waldman, *Introduction: Framing Regulatory Managerialism as an Object of Study and Strategic Displacement*, 86 L. & CONTEMP. PROBS. i, ix-x (2023); *see generally*

values largely in mind, we have explained elsewhere, can lead to a "hands-off deference" to values choices[27] and a "skewing of public legal norms by private interests."[28] Risk assessment, "a key technique of managerialism directed at formalizing and constraining agency decision-making,"[29] was developed as "a political technology intended to discipline agencies, rather than a tool for revealing truths about the world."[30] Risk-based regimes, accordingly, facilitate a "light-touch"[31] approach to governance, while the centering of public values demands instead "activist" and "dynamic" regulators.[32]

In the end, managerial approaches to technological "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."[33] In this way, as Kaminski describes, a risk-based metric for AI governance alone fails to account for "dignitary and justificatory concerns about algorithmic decision-making."[34] A legitimate AI governance regime must account in a thick manner for the ways in which it identifies and protects the rights and public values that are at risk.

At the same time, while we have advocated for the importance of privileging human and public rights in technology governance, we have also raised cautions about the method for doing so. In particular, we have pointed to the "range of human rights and other public values"[35]—some of which might be in tension—that could be embedded in technology design. The choice between them, we have argued, must derive from democratically- and

Frank A. Pasquale, *Power and Knowledge in Policy Evaluation: From Managing Budgets to Analyzing Scenarios*, 86 L. & CONTEMP. PROBS. 39, 43 n.20 (2023) ("There is, as Robert Post has observed, a critical distinction between governance and management." (citing Robert Post, *Between Governance and Management: The History and Theory of the Public Forum*, 34 UCLA L. REV. 1713, 1788 (1986))).

27. Bamberger, *Technologies of Compliance*, *supra* note 3, at 684.

28. *Id.* at 726.

29. William Boyd, *With Regard for Persons*, 86 L. & CONTEMP. PROBS. 101, 104 (2023).

30. *Id.*

31. Cohen & Waldman, *supra* note 26, at xv.

32. Bamberger, *Technologies of Compliance*, *supra* note 3, at 735; *see also* KENNETH A. BAMBERGER & DEIRDRE K. MULLIGAN, PRIVACY ON THE GROUND 187, 225 (2015) (documenting the importance of such "activist" regulators in outcomes-based governance regimes).

33. Deirdre K. Mulligan & Kenneth A. Bamberger, *Allocating Responsibility in Content Moderation: A Functional Framework*, 36 BERKELEY TECH. L.J. 1091, 1095 (2021).

34. Margot E. Kaminski, *Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability*, 92 S. CAL. L. REV. 1529, 1533 (2019).

35. Mulligan & Bamberger, *Governance-By-Design*, *supra* note 2, at 702 (noting that "we lack a comprehensive approach—a doctrine, a set of metrics, as well as tools—for resolving design wars while accounting for the range of human rights and other public values").

deliberatively-legitimate engagement rather than through "design wars."[36] We should moreover, be cautious about "'baking' human and public rights values into technology systems" at any one given moment, "because of the strength and durability of a decision to govern by design."[37] And when a technical component—for example a model or data set—may be used in a wide range of technical systems, and in a wide range of contexts, sound governance requires attention to the diversity of values configurations that may be required to meet regulatory goals and normative expectations. Thus, policymakers should strive for policies that "steer the protection of rights and values to the least intrusive point" to "enable the promotion of values rather than fixing them in determinatively," and provide "technological hooks that permit different value choices in different contexts."[38]

B. INSTITUTIONAL DESIGN IN AI REGULATION: WHO SHOULD GOVERN?

Concerns about the fora and processes through which choices are made about what rights are prioritized in AI governance directly implicate questions of institutional design. A key component of institutional design is ensuring the appropriate expertise necessary for governance.

Accordingly, a second debate in AI governance concerns whether AI should be regulated through new, rather than existing, institutions. This debate came to the fore during a series of hearings in mid-2023 convened by the Senate Judiciary Committee, together with the Senate Homeland Security and Governmental Affairs Committee, during which senators spoke with AI researchers and industry leaders to discuss whether new developments in AI technology warrant additional regulation.[39] Many AI industry leaders, including Sam Altman, the CEO of OpenAI, Jared Kaplan and Jack Clark, the co-founders of Anthropic AI, and Elon Musk, the CEO of Tesla and X, also participated in a series of meetings convened by the Bipartisan Senate AI

36. *Id.*; *see also* Deirdre K. Mulligan & Kenneth A. Bamberger, *Apple v. FBI: Just One Battle in the 'Design Wars'*, LAW.COM (Mar. 21, 2016), https://www.law.com/sites/lawcomcontrib/2016/03/18/apple-v-fbi-just-one-battle-in-the-design-wars./.

37. Mulligan & Bamberger, *Governance-By-Design*, *supra* note 2, at 750.

38. *Id.*

39. *See, e.g.*, Michael D. Bopp, Roscoe Jones Jr., Alexander Southwell, Amanda H. Neely, Daniel P. Smith, Frances Waldmann, Kirsten Bleiweiss & Madelyn Mae La France, *"Oversight of AI: Rules for Artificial Intelligence" and "Artificial Intelligence in Government" Hearings*, GIBSON DUNN (June 6, 2023), https://www.gibsondunn.com/oversight-of-ai-rules-for-artificial-intelligence-and-artificial-intelligence-in-government-hearings/ (describing takeaways from Senate Judiciary Hearings on AI).

Working Group, led by Senate Majority Leader Chuck Schumer, Senator Mike Rounds, Senator Martin Heinrich, and Senator Todd Young in 2023.[40]

During these hearings and discussions, industry leaders supported the creation of new institutions to regulate AI.[41] At Senate hearings convened on May 16, 2023, for example, Sam Altman of OpenAI endorsed the formation of a "new agency [for AI] that licenses any effort above a certain scale of capabilities and can take that license away and ensure compliance with safety standards."[42] When pressed by Senator Lindsey Graham on whether the most effective way to combat security threats stemming from AI would be to "have an agency that is more nimble and smarter than Congress" overseeing AI, Altman replied that OpenAI would be "enthusiastic" about the creation of such an agency.[43] When Christina Montgomery, IBM's chief privacy and trust officer, raised doubts about whether a new agency was necessary for effective regulation of AI or whether existing institutions were sufficient, she was quickly shut down by Professor Marcus, Senator Graham, and others at the hearing who were apparently already convinced that a new federal agency to regulate AI was the right path forward.[44] Industry representatives such as Eric Schmidt, the former CEO of Google, and Mustafa Suleyman, the CEO of Microsoft AI, have also supported the creation of nonregulatory expert-led bodies to inform governments about developments in the AI space, comparable to the Intergovernmental Panel on Climate Change (IPCC).[45]

---

40. MAJORITY LEADER CHUCK SCHUMER, SEN. MIKE ROUNDS, SEN. MARTIN HEINRICH & SEN. TODD YOUNG, BIPARTISAN SENATE AI WORKING GROUP, DRIVING U.S. INNOVATION IN ARTIFICIAL INTELLIGENCE: A ROADMAP FOR ARTIFICIAL INTELLIGENCE POLICY IN THE UNITED STATES SENATE (May 2024), https:// https://www.schumer.senate.gov/imo/media/doc/Roadmap_Electronic1.32pm.pdf.

41. *See* Alexander C. Kurtz, *Regulating into the Void: Existential Uncertainty from A.I. Necessitates a New Federal Research Agency*, B.C. INTELL. PROP. & TECH. F. 1 (2024).

42. *Oversight of A.I.: Rules for Artificial Intelligence: Hearing Before the Subcomm. on Priv., Tech., and the L.*, 118th Cong. (2023), https://www.govinfo.gov/content/pkg/CHRG-118shrg52706/html/CHRG-118shrg52706.htm.

43. *Id.*

44. *Id.*

45. Mustafa Suleyman & Eric Schmidt, *Mustafa Suleyman and Eric Schmidt: We Need an AI Equivalent of the IPCC*, FIN. TIMES (Oct. 18, 2023), https://www.ft.com/content/d84e91d0-ac74-4946-a21f-5f82eb4f1d2d; Mustafa Suleyman, Mariano-Florentino (Tino) Cuéllar, Ian Bremmer, Jason Matheny, Philip Zelikow, Eric Schmidt & Dario Amodei, *Proposal for an International Panel on Artificial Intelligence (AI) Safety (IPAIS): Summary*, CARNEGIE ENDOWMENT FOR INT'L PEACE (Oct. 27, 2023), https://carnegieendowment.org/posts/2023/10/proposal-for-an-international-panel-on-artificial-intelligence-ai-safety-ipais-summary?lang=en (proposing an International Panel on Artificial Intelligence Safety inspired by the Intergovernmental Panel on Climate Change, a United Nations body that assesses the scientific, technical, and socio-economic information relevant to understanding climate change, its impacts, and potential risks).

Although the academic literature on AI institutions is still in its infancy, some scholars have echoed industry leaders' and policymakers' calls for the creation of a new federal agency to regulate AI in the United States, at least vis-à-vis "existential risks" such as mass access to bioweapons facilitated by generative AI (GAI) or other evolutions of AI that could be "misaligned" with human interests.[46] For example, during the same Senate hearing at which Sam Altman of OpenAI testified, Professor Gary Marcus, an expert on AI policy based at New York University, voiced a similar view to Altman's. Marcus claimed that a new, "Cabinet-level organization" with technical expertise should be propped up to regulate AI at the federal level.[47]

These calls from industry leaders and academics for new agencies to regulate AI gained traction in Congress. In January 2024, Senators Michel Bennet, Elizabeth Warren, Lindsey Graham, and Peter Welch wrote a letter to Senator Schumer requesting the creation of a new federal agency to regulate digital markets including AI.[48] The Senators claimed that the hearings made evident the need to meet "the transformative challenge of AI with a thoughtful and effective regulatory framework," and voiced their belief that "this moment [in AI development] requires a new federal agency to protect consumers, promote competition, and defend the public interest."[49] The letter cited other instances—including the creation of the Food and Drug Administration in 1906 and the creation of the Federal Communications Commission in 1934—

46. *See, e.g.*, Bryan Druzin, Anatole Boute & Michael Ramsden, *Confronting Catastrophic Risk: The International Obligation to Regulate Artificial Intelligence*, 46 MICH. J. INT'L L. 173, 182, 189, 198 (2025) (arguing that "the precautionary principle requires states to act, as waiting for conclusive scientific evidence before addressing its potential existential risk may prove too late.").

47. *Oversight of A.I.: Rules for Artificial Intelligence: Hearing Before the Subcomm. on Priv., Techn., and the L.*, *supra* note 42. Industry leaders have changed their position, arguing for "light-weight" regulations at the federal level to limit the proliferation of state laws and no longer calling for new agencies. *See, e.g.*, *Winning the AI Race: Strengthening U.S. Capabilities in Computing and Innovation*, *Hearing Before the S. Comm. on Com., Sci., & Transp.*, 119th Cong. (May 8, 2025), https://www.govinfo.gov/content/pkg/CHRG-119shrg61426/pdf/CHRG-119shrg61426.pdf, (responses of Sam Altman, CEO, OpenAI ("One federal framework that is light touch that we can understand and that lets us, you know, move with the speed that this moment calls for seems important and fine, but the sort of every state takes a different approach here, I think would be quite burdensome and significantly impair our ability to do what we need to do."), and Brad Smith, CEO, Microsoft ("[T]he United States needs to be in the game internationally to influence the rest of the world. And you cannot be in the game if you do nothing. You must do something. So you take . . . a lightweight approach . . . and then you build support around it.")).

48. Letter from Senators Michael F. Bennet, Lindsey O. Graham, Elizabeth Warren & Peter Welch to Chuck Schumer, Majority Leader (Jan. 23, 2024), https://www.warren.senate.gov/imo/media/doc/F46611AE0DF77719F8B18AE6C197C52B.joint-letter-to-schumer-on-digital-platform-agency.pdf.

49. *Id.* at 1.

where Congress, confronted with the "emergence of complex, risk-prone industries . . . elected to create [new] regulatory bodies."[50] And the Senators justified their call for a new agency by claiming that "[p]arceling out oversight to various agencies will result in a fragmented regulatory landscape ripe for exploitation by companies with market caps greater than many countries' gross domestic product."[51]

Legal scholars and policymakers have long explored the benefits of creating new, specialized agencies to deal with novel problems,[52] such as the ability of such institutions to develop expertise to bear on topics that existing agencies do not have the knowledge or ability to handle,[53] or to coordinate action among a wide range of stakeholders at the state and federal level. Such arguments about expertise underpinned the establishment of the FAA in 1958,[54] while those about coordination provided the primary rationale behind the creation of the EPA.[55]

Yet many of the regulatory concerns surrounding AI are distinct from those that led to the founding of the FAA or the EPA. Whereas the invention of airplanes resulted in the advent of an entirely new sector, the same cannot be said of AI. AI has many uses and will be embedded in products and services across various sectors, but it is not a *thing* in and of itself. By nature, AI will have countless different applications and will implicate a range of rights and values.

The risk of regulatory capture[56] of specialist agencies by the interest group they regulate, moreover, is especially pronounced[57] in the AI context, because of the limited number of powerful players in the space, all of whom have nearly endless financial resources and similar regulatory interests. Compared to a generalist institution like the Office of Information and Regulatory Affairs or even an agency that regulates various industries like the EPA, the benefits of

50. *Id.* at 2.

51. *Id.*

52. *See generally* Rachel E. Barkow, *Insulating Agencies: Avoiding Capture Through Institutional Design*, 89 TEX. L. REV. 15 (2010).

53. *See id.* at 20.

54. *See generally* John W. Gelder, *Air Law - The Federal Aviation Act of 1958*, 57 MICH. L. REV. 1214 (1959).

55. *See* Jonathan H. Adler, *The Environmental Protection Agency Turns Fifty*, 70 CASE W. RSRV. L. REV. 871 (2020).

56. Capture being the phenomenon by which "organized interest groups successfully act to vindicate their goals through government policy at the expense of the public interest." Michael A. Livermore & Richard L. Revesz, *Regulatory Review, Capture, and Agency Inaction*, 101 GEO. L.J. 1337, 1340 (2013).

57. *See* Jonathan R. Macey, *Organizational Design and Political Control of Administrative Agencies*, 8 J. L. ECON. & ORG. 93, 99 (1992) ("The interest group that is regulated by a single regulatory agency will be able to influence that agency to a far greater extent than the interest groups that must 'share' their agency with a variety of other interest groups.").

capturing a specialized institution are comparatively higher for industry players. They have the undivided attention of regulators vis-à-vis a given regulation, they do not have to compete with other industries or interests for regulators' attention, and they have incentives to coordinate when their interests are aligned.[58] The Senators correctly identified the challenges posed by the deep pockets of industry, but the creation of a new entity to regulate a small number of companies with such wealth creates the perfect conditions for regulatory capture.

Finally, and perhaps most importantly in light of the analysis below, the assignment of AI regulation to a specialized regulator would create its own set of problems of legitimacy and expertise. For as we discuss, *infra*, the impact of AI on real interests and real individuals does not play out at a theoretical level, but on the ground, in concrete contexts that are already often governed by regulatory bodies with domain competence. While these organs of government might face challenges of technological expertise, they are already invested with legitimacy in administering laws and protecting and enforcing rights independently defined in legislation and regulation. And they possess critical expertise relevant to the ways that AI would affect rights in particular contexts. Importantly, the fields these agencies regulate have their own logics—epistemological and ethical—that control how they produce and act upon knowledge. Redirecting regulatory authority to a specialized agency would decenter that domain expertise in AI governance. Centering the tool or method—AI—rather than the domain creates the conditions for regulatory displacement, or a form of regulatory arbitrage by which firms might seek to undermine, escape, or preclude meaningful domain-specific regulation by appeals to generalized, and less tailored or exacting, mandates.

These concerns suggest that addressing regulatory expertise requires a fundamentally different approach—one that better prioritizes public values. This approach should treat AI as "normal technology"[59] that can be shaped towards different ends,[60] rather than fetishizing it as uniquely ungovernable. Effective AI regulation must be domain-specific, vindicate relevant rights, and ensure meaningful participation by affected stakeholders.

---

58. *See generally* Livermore & Revesz, *supra* note 56.

59. Arvind Narayanan & Sayash Kapoor, *AI as Normal Technology*, KNIGHT FIRST AMEND. INST. (Apr. 15, 2025), https://knightcolumbia.org/content/ai-as-normal-technology.

60. Langdon Winner, *Do Artifacts Have Politics?*, 109 DAEDALUS 121, 123 (1980) (describing that the politics of a technical system can arise through their design processes or for a narrow group of technologies through inherent properties). Narayanan and Kapoor's "normal technology" argument, in part, positions AI in the first category—it is pliable and configurable and can support different social and political arrangements.

III. OUR FRAMEWORK FOR RECENTERING PUBLIC VALUES IN TECHNOLOGY GOVERNANCE: AN ALTERNATIVE TO THE AI DEBATES

The framework for technology governance set forth in our earlier work[61] reflects the concerns articulated by different sides in the prevailing debates over AI regulation. AI governance requires choices about, a focus on, and a comprehension of, the public rights implicated by the technology. It requires expertise sufficient to comprehend the risks posed to those rights in concrete contexts. Yet our regulatory principles suggest an alternate approach that accounts for both the limits of the risk/rights binary and the shortcomings of traditional legal and governance bodies in the face of technology challenges. In particular, they address challenges that include: inadequate technical expertise; the difficulty of accounting for multiple rights and values, or contests between them; and the absence of fora for, and meaningful assessments to inform, discussions regarding how to prioritize values in the context of technological development and implementation. Looking forward, these principles seek to ensure governance processes that prioritize rights and reflect fundamental democratic norms, are free from capture or caprice, and avoid technological fixes that privilege singular values and often disfavor human rights.

To do so, our framework sets forth three interrelated regulatory principles for centering the "public" in technology governance.

First, such governance must *privilege human and public rights*, meaning on the one hand that design must be oriented intentionally with rights in mind and, on the other, that it should reflect the reality that a multiplicity of rights might be impacted by AI systems, and that the rights impacted might depend on domain and context. Accordingly, design should allow, wherever possible, for flexibility in deployments and make values choices visible and configurable for deployers and users. This principle does not, however, reject a regulatory focus on risk or risk assessment. Rather, it requires that risk management must be responsive to rights in context and reflect the fact that the rights we care about and how we evaluate and mitigate risks to them vary by domain.

Second, AI governance requires *agencies with appropriate domain competence and sociotechnical expertise*. These are venues in which "government and other stakeholder groups have deep access to technical expertise"[62] in addition to "domain-specific expertise."[63] While it focuses on institutional design, this principle points away from debates over the specific structure of regulatory institutions to the question of whether regulators possess the right tools to

61. *See* discussion *supra* notes 2–10 and accompanying text.
62. Mulligan & Bamberger, *Governance-By-Design*, *supra* note 2, at 759.
63. Mulligan & Bamberger, *Procurement as Policy*, *supra* note 1, at 837.

avoid goal-myopia, foster the consideration of competing values, claim legitimacy in the choice between values, and understand the implications for them of technical decisions.[64] Such tools are required in terms of both authority and competence, including technical—or our preferred term, sociotechnical—expertise.

Finally, our third regulatory principle mandates *maintaining the publicness of policymaking*, specifically by employing "mechanisms that translate traditional commitments to participation and transparency to the technology context, in ways that address the intricate way in which policy is embedded in technical design and implementation choices."[65]

Stakeholder-participation and community engagement processes go part of the way to integrating public deliberation of values in technology policymaking. Yet, as we have argued,[66] they cannot alone address the challenge that technology design presents for governance principles of participation and transparency because of the difficulty in comprehending the value implications of technology systems from the outside. The challenge is two-fold. On the one hand, those systems' opacity obscures the values embedded in technical infrastructure from the start; on the other, the impact on rights and values plays out not just at the moment of initial design, but "in a continuum—at design time, configuration time, and run time."[67] Thus the impact on rights and interests varies throughout these phases in different contexts and applications.

Accordingly, we contend: (1) meaningful participation requires sociotechnical expertise among both regulators and stakeholders; and (2) meaningful transparency must involve "political visibility" into the existence and political nature of questions being resolved during design and use.[68] To that end, we have argued for the use of values-surfacing tools in technical design, drawing on a range of approaches that provide clarity over the properties embedded in code and other technical artifacts, as well as its performance.

A wide range of tools can keep values in view at different stages of development, configuration, and deployment. For example, formal methods provide clarity about a technology's properties during development and use. Impact assessments and data and system documentation can serve as *boundary*

64. Mulligan & Bamberger, *Governance-By-Design*, *supra* note 2, at 759–69.

65. *Id.* at 770.

66. *Id.* at 770–83.

67. *Id.* at 773 (citing David D. Clark, John Wroclawski, Karen R. Sollins & Robert Braden, *Tussle in Cyberspace: Defining Tomorrow's Internet*, 13 IEEE/ACM TRANSACTIONS ON NETWORKING 462, 463 (2005)).

68. *Id.* at 776.

objects[69] that facilitate designers and the publics deliberation on the "technocratic and democratic elements"[70] of systems. These tools can support learning, communication and deliberation across diverse stakeholders on a shared endeavor such as the design of a sociotechnical system. For example, the Census Bureau published data artifacts (aka "demonstration data") produced by different system implementations of differential privacy to scaffold public understanding of those embedded policy choices.[71] This demonstration data served as a boundary object "allow[ing] stakeholders to interactively and intuitively explore the impact of potential implementation choices on their equities."[72]

As we discuss, *infra*,[73] identifying the methods and the metrics for meaningfully aligning a sociotechnical system with relevant values and assessing its impact on public rights and values presents challenges. Yet such tools and methods offer the capacity to both catalyze deliberation about the technical aspects of system design and also to surface the political implications of those choices. These tools and methods thus "bridge the dual deliberation requirements of substantive expertise and political visibility"[74] by creating "different frameworks and bring new considerations to bear in agency actions."[75] And at the same time, they "bridge the gulf between the substantive domain expertise of agency staff and the frameworks and knowledge of outside experts,"[76] facilitating "participation by issue experts and by stakeholders who might otherwise be unaware of relevant risks and technological alternatives."[77]

69. Susan Leigh Star & James R. Griesemer, *Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39*, 19 SOC. ST. SCI. 387, 388 (1989).

70. Mulligan & Bamberger, *Procurement as Policy*, *supra* note 1, at 842.

71. Amina A. Abdu, Lauren M. Chambers, Deirdre K. Mulligan & Abigail Z. Jacobs, *Algorithmic Transparency and Participation Through the Handoff Lens: Lessons Learned from the U.S. Census Bureau's Adoption of Differential Privacy*, FACCT '24: PROCS. OF THE 2024 ACM CONF. ON FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY 1150–62 (2024).

72. *Id.* at 1158.

73. *See infra* Part IV.

74. Mulligan & Bamberger, *Procurement as Policy*, *supra* note 1, at 842.

75. *Id.* at 844.

76. *Id.*

77. Mulligan & Bamberger, *Governance-By-Design*, *supra* note 2, at 765.

IV. REFLECTING OUR GOVERNANCE PRINCIPLES: EXAMPLES FROM THE BIDEN-HARRIS ADMINISTRATION'S APPROACH TO AI GOVERNANCE

The Biden-Harris Administration moved forward with a suite of AI Governance initiatives that reflect these governance principles.[78] We focus on three examples.

First, the Administration made clear that AI and automated decision-making systems—regardless of who built or used them—should privilege the public's rights—particularly human rights—and safety.[79] As the President's Science and Technology Advisor and Director of OSTP, Arati Prabhakar said: "[w]e're in choppy waters with this rapidly changing technology, and that

78. *See* discussion *infra* Part IV. While Mulligan served as Principal Deputy U.S. Chief Technical Officer in the White House Office of Science and Technology Policy (OSTP) in the Biden-Harris Administration during a key period of AI policy development, this is a retrospective analysis of some, but not all, of the Administration's key actions, not a claim that our research framing specifically drove its approach. In addition, this review does not include some key Administration actions designed to limit adversarial uses of AI by foreign actors including export controls on the most powerful models, the "Diffusion Rule," Framework for Artificial Intelligence Diffusion, 90 Fed. Reg. 4544 (Jan. 15, 2025), which revised the Export Administration Regulations' (EAR) controls on advanced computing integrated circuits (ICs) and added controls on AI model weights for certain advanced closed-weight dual-use AI models to protect U.S. national security and foreign policy interests among other things, rescinded three days before its effective date on May 12, 2025. *See Department of Commerce Rescinds Biden-Era Artificial Intelligence Diffusion Rule, Strengthens Chip-Related Export Controls*, BUR. OF INDUS. & SEC., U.S. DEP'T OF COM. (May 12, 2025), https://media.bis.gov/sites/default/files/documents/05.07%20Recission%20of%20AI%20Diffusion%20Press%20Release-2.pdf; and the "BIS Rule," issued by the Department of Commerce Bureau of Industry and Standards pursuant to the Defense Production Act of 1950 (DPA), which requires quarterly reporting from companies developing or demonstrating an intent to develop dual-use foundation AI models and those with large-scale computing clusters for AI model training runs over 10^26 computation operations or acquiring/possessing a computer cluster with data center networking over 300 Gbits. Establishment of Reporting Requirements for the Development of Advanced Artificial Intelligence Models and Computing Clusters, 89 Fed. Reg. 73612 (proposed Sep. 11, 2024) (to be codified at 15 C.F.R. pt. 702). This rule has not yet been finalized, and it seems unlikely that it will be given recent Administration actions. During the Biden-Harris Administration, BIS required several AI companies to make such disclosures pursuant to E.O. 14110.

79. Exec. Order. No. 14,110 § 2, 88 Fed. Reg. 75191, at 75191–93 (Nov. 1, 2023) (signed Oct. 30, 2023) (setting out the principles and policies to guide AI including: § 2(a) ensuring that AI is safe and secure; § 2(b) promoting responsible innovation, competition and collaboration; § 2(c) supporting workers; § 2(d) advancing equity and civil rights; § 2(e) protecting consumers; § 2(f) protecting privacy and civil liberties.; § 2(g) building the necessary expertise—technical, managerial, ethical, legal, etc.—in government; § 2(h) leading global development and adoption so that AI "benefits the whole world, rather than exacerbating inequities, threatening human rights, and causing other harms"). This Executive Order was revoked by the Trump administration on January 20, 2025.

means it's more important than ever to steer by the light of these fundamental values."[80] At the first AI summit, Vice President Harris stated that:

> [T]he urgency of this moment must then compel us to create a collective vision of what this future must be. A future where AI is used to advance human rights and human dignity, where privacy is protected and people have equal access to opportunity, where we make our democracies stronger and our world safer.[81]

Harris further stated that the Administration "believe[d] that all leaders from government, civil society, and the private sector have a moral, ethical, and societal duty to make sure that AI is adopted and advanced in a way that protects the public from potential harm and that ensures that everyone is able to enjoy its benefits."[82] Even in the context of national security, the President centered rights and values, writing that:

> Success for the United States in the age of AI will be measured not only by the preeminence of United States technology and innovation, but also by the United States' leadership in developing effective global norms and engaging in institutions rooted in international law, human rights, civil rights, and democratic values.[83]

To protect the public's rights and safety the Administration took a layered approach to governance, empowering domain specific regulators and creating a new set of expertise to develop risk identification and mitigation techniques.

Second, the Administration expanded the AI and AI-enabling expertise within the government; clarified the importance of civil rights, privacy civil liberties, and other rights-centered experts in AI design and use; and created opportunities for AI experts in government, academia, civil society, and the corporate sector to collaboratively build a body of knowledge and practice to support the identification and development of testing, evaluation, validation

80. OSTP Director Arati Prabhakar Remarks on Managing AI's Risks to Seize its Benefits, as Prepared for Delivery at the *Carnegie Endowment for International Peace*, WHITE HOUSE (Nov. 14, 2023), https://bidenwhitehouse.archives.gov/ostp/news-updates/2023/11/14/remarks-of-arati-prabhakar-at-carnegie-endowment-for-international-peace/.

81. Vice President Kamala Harris, Remarks on the Future of Artificial Intelligence at the U.S. Embassy, London, United Kingdom (Nov. 1, 2023), https://bidenwhitehouse.archives.gov/briefing-room/speeches-remarks/2023/11/01/remarks-by-vice-president-harris-on-the-future-of-artificial-intelligence-london-united-kingdom/.

82. *Id.*

83. Memorandum on Advancing the United States' Leadership in Artificial Intelligence; Harnessing Artificial Intelligence to Fulfill National Security Objectives; and Fostering the Safety, Security, and Trustworthiness of Artificial Intelligence, 2024 DAILY COMP. PRES. DOC. 00945, 3 (Oct. 24, 2024).

and risk management practices to support federal agencies and others developing and using, as well as regulating, AI systems.[84]

Third, the Administration instituted practices to expose and interrogate the values embedded in federal agency AI *use cases* throughout design and deployment, and scaffold public participation in their design, risk mitigation activities, and evaluations.[85]

Together, these initiatives reflect the interdependent principles of our governance-by-design framework, chart a new path in AI governance, and diverge from many of the dominant AI debates. The AI governance framework is not a traditional risk management framework but rather layers new knowledge onto and offers new expertise to support sectoral regulations and regulators that define rights and determine the level of tolerable risk within domains. It offers a model by which AI governance can be both rights and risk based, and can leverage new and existing institutions effectively, all while engaging key stakeholders from the public and private sectors, civil society, and academia, at every stage from AI testing to deployment.

A. PRIVILEGING HUMAN AND PUBLIC RIGHTS: MAINTAINING EXPERT AGENCY AUTHORITY WHILE BUILDING A SHARED KNOWLEDGE BASE FOR RISK ASSESSMENT METHODS AND PRACTICES

As a foundational matter, the Biden-Harris Administration made it clear from the outset that, regardless of the parties involved in their development and utilization, AI and automated decision-making systems must be developed and implemented in ways that account for generalized and sector-specific systemic risks, but that the fundamental priority should be mitigating risks to public rights, with a particular emphasis on human rights and safety.[86]

The Administration issued the Blueprint for an AI Bill of Rights (AI BoR), a clear statement affirming the centrality of the public's rights and safety in the design and use of AI and articulating processes and practices to protect them.[87] The White House OSTP published the AI BoR in October 2022, before

84. Discussed, *infra*, at Section IV(B).

85. Discussed, *infra*, at Section IV(C).

86. Below we focus on a subset of the Administration's AI actions to illustrate the focus on rights and safety. Other actions to limit risk include the AI diffusion rule, subsequently withdrawn under the Trump Administration, and the BIS reporting rule. Framework for Artificial Intelligence Diffusion, 90 Fed. Reg. 4544 (Jan. 15, 2025); *Department of Commerce Announces Rescission of Biden-Era Artificial Intelligence Diffusion Rule, Strengthens Chip-Related Export Controls*, BUR. OF INDUS. & SEC., U.S. DEP'T OF COM. (May 13, 2025), https://www.bis.gov/press-release/department-commerce-announces-rescission-biden-era-artificial-intelligence-diffusion-rule-strengthens.

87. WHITE HOUSE OFF. OF SCI. & TECH. POL'Y, BLUEPRINT FOR AN AI BILL OF RIGHTS: MAKING AUTOMATED SYSTEMS WORK FOR THE AMERICAN PEOPLE (2022), https://bidenwhitehouse.archives.gov/ostp/ai-bill-of-rights/ [hereinafter AI BOR].

ChatGPT and other large language models garnered the attention of the media, the public, and policymakers.[88] That document set out a framework to guide the design, development, and deployment of automated systems so that they protect the rights of the American public and reinforce our nation's longstanding values. The AI BoR announced five principles to guide the design, use, and deployment of automated systems.[89] Automated decision-making systems (ADS), including AI systems, should be safe and effective, free from algorithmic discrimination, and respect data privacy.[90] Individuals should know when ADS are being used and receive an explanation of how it impacts decisions about them.[91] Lastly, individuals should be able to opt for a human alternative rather than an ADS process and have easy access to a person who can quickly consider and address problems with ADS systems.[92] The AI BoR pays special attention to domains, including criminal justice and education, where automated systems can have significant adverse effects on human rights, civil liberties, and civil rights.[93] It calls for limitations on surveillance, including stating that continuous surveillance and monitoring should not be used in settings where it is likely to limit rights, opportunities, or access.[94]

The Administration quickly moved from rights-based principles to commitments to protection. First, in February 2023, President Biden directed the federal government to root out bias in the design and use of new technologies, such as artificial intelligence; to protect the public from algorithmic discrimination; and to bring civil rights offices into conversations about the design, procurement, and use of automated systems.[95]

Second, the Administration pushed the private sector to protect the public's rights and safety during AI development and use. In July 2023, President Biden garnered voluntary commitments from leading AI companies, including Amazon, Anthropic, Google, Inflection, Meta, Microsoft, and OpenAI, to engage in a set of practices designed to identify and mitigate risks to the public's rights and safety.[96] Specifically, these companies committed to

88. *See When Was ChatGPT Released*, SCRIBBR, https://www.scribbr.com/frequently-asked-questions/when-was-chatgpt-released/ (last visited Sep. 16, 2025) (noting that Chat GPT was publicly released on November 30, 2022, and that at the time of its release, was described as a "research preview").

89. AI BOR, *supra* note 87, at 5–7.

90. *Id.* at 5–6.

91. *Id.* at 6.

92. *Id.* at 7.

93. *See, e.g.*, *id.* at 36.

94. *Id.* at 6, 30, 34.

95. Exec. Order No. 14,091, 88 Fed. Reg. 10825 (Feb. 16, 2023).

96. *FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI*, WHITE HOUSE (July 21, 2023), https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2023/07/21/

"internal and external security testing of their AI systems before their release"; "sharing information across the industry and with governments, civil society, and academia on managing AI risks"; and "prioritizing research on the societal risks that AI systems can pose, including on avoiding harmful bias and discrimination, and protecting privacy," among other commitments.[97]

Third, in October 2023, President Biden issued an Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence (EO 14110).[98] This executive order directed sweeping actions across the government. It opens by affirming the values—privacy, equity, etc.—that AI design and use must respect and mitigate its impact on (including by specifically stating that "Americans' privacy and civil liberties must be protected as AI continues advancing").[99] EO 14110 specifically notes that while the Administration aimed to "promote responsible uses of AI that protect consumers, raise the quality of goods and services, lower their prices, or expand selection and availability," it would also ensure protections for important rights "especially important in critical fields like healthcare, financial services, education, housing, law, and transportation, where mistakes by or misuse of AI could harm patients, cost consumers or small businesses, or jeopardize safety or rights."[100] Through EO 14110, the Administration encouraged existing agencies to make full use of their authorities to address domain-relevant AI risks and, as we discuss, *infra*, set up a hiring surge and a new institution to expand the AI and AI-enabling expertise available across federal agencies.

In March and July 2024, respectively, the Office of Management and Budget (OMB) issued guidance to federal agencies on the responsible use of AI.[101] In October 2024, the White House issued a national security

---

fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/.

97. *FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Eight Additional Artificial Intelligence Companies to Manage the Risks Posed by AI*, WHITE HOUSE (Sep. 12, 2023), https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2023/09/12/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-eight-additional-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/#:~:text=The%20companies%20commit%20to%20internal,circumvent%20safeguards%2C%20and%20technical%20collaboration.

98. Exec. Order No. 14,110, 88 Fed. Reg. 75191 (Nov. 1, 2023) (signed Oct. 30, 2023).

99. Exec. Order No. 14,110 § 1(f), 88 Fed. Reg. at 75193.

100. Exec. Order No. 14,110 § 1(e), 88 Fed. Reg. at 75193.

101. OFF. OF MGMT. & BUDGET, EXEC. OFF. OF THE PRESIDENT, MEMO. NO. M-24-10, ADVANCING GOVERNANCE, INNOVATION, AND RISK MANAGEMENT FOR AGENCY USE OF ARTIFICIAL INTELLIGENCE (Mar. 28, 2024), https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf [hereinafter OMB MEMO. M-24-10]. Executive Order 14110 directed OMB to fulfill unmet obligations under 40 U.S.C. § 11301 (the AI in Government Act and the Advancing American AI Act).

memorandum (NSM) and accompanying framework establishing similar requirements for agencies on use of AI on national security systems.[102] The OMB guidance directs agencies to ensure the public's rights and interests—including privacy, nondiscrimination and equity, security, and accessibility—are protected and risks to them are mitigated in system design and use.[103] OMB's guidance to agencies established the most specific and rigorous set of requirements for the design and use of technology by government agencies. The NSM is similarly robust and includes prohibitions on certain uses of AI.[104] OMB subsequently developed AI specific procurement policy to ensure agencies received information and secured the ability to interact with AI systems necessary to evaluate the impact of federal AI use cases on the public's rights and safety.[105]

The Biden-Harris Administration affirmed the centrality of human rights and democratic values in multinational AI policy as well. In November 2023, Vice President Kamala Harris participated in a Global Summit on AI Safety hosted by former Prime Minister Rishi Sunak of the United Kingdom.[106] Prior to this engagement, Vice President Harris laid out the Administration's vision of a future where "AI is used to advance human rights and human dignity, where privacy is protected and people have equal access to opportunity,"[107] and announced the Administration's joint commitment with thirty other countries on the responsible use of military AI.[108] In her speech, she pushed back against an exclusive emphasis on AI risk related to chemical, biological, radiological and nuclear risks, stating:

> But let us be clear. There are additional threats that also demand our action—threats that are currently causing harm and which, to many people, also feel existential. Consider, for example: When a senior is kicked off his healthcare plan because of a faulty AI algorithm, is

102. WHITE HOUSE, FRAMEWORK TO ADVANCE AI GOVERNANCE AND RISK MANAGEMENT IN NATIONAL SECURITY 1–2 (Oct. 24, 2024).

103. OMB MEMO. M-24-10, *supra* note 101.

104. 2024 DAILY COMP. PRES. DOC. 00945, *supra* note 83; WHITE HOUSE, *supra* note 102.

105. OFF. OF MGMT. & BUDGET, EXEC. OFF. OF THE PRESIDENT, MEMO. NO. M-24-18, ADVANCING THE RESPONSIBLE ACQUISITION OF ARTIFICIAL INTELLIGENCE IN GOVERNMENT (Sep. 24, 2024), https://www.whitehouse.gov/wp-content/uploads/2024/10/M-24-18-AI-Acquisition-Memorandum.pdf [hereinafter OMB MEMO. M-24-18].

106. Press Release, Statement by Press Secretary Kirsten Allen on the Vice President's and Second Gentleman's Travel to the United Kingdom (Oct. 26, 2023) (stating that the Vice President would deliver a speech on AI on November 1 in London and represent the United States at the Global Summit on AI Safety at Bletchley Park on November 2), https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2023/10/26/statement-by-press-secretary-kirsten-allen-on-the-vice-presidents-and-second-gentlemans-travel-to-the-united-kingdom/.

107. Vice President Kamala Harris, *supra* note 81.

108. *Id.*

> that not existential for him? When a woman is threatened by an abusive partner with explicit, deep-fake photographs, is that not existential for her? When a young father is wrongfully imprisoned because of biased AI facial recognition, is that not existential for his family? And when people around the world cannot discern fact from fiction because of a flood of AI-enabled mis- and disinformation, I ask, is that not existential for democracy? Accordingly, to define AI safety, I offer that we must consider and address the full spectrum of AI risk—threats to humanity as a whole, as well as threats to individuals, communities, to our institutions, and to our most vulnerable populations.[109]

The Administration ensured that international conversations—from those in the UN General Assembly to those held by the G7—centered human rights. The United States, for example, led the drafting and adoption of the *United Nations General Assembly resolution on trustworthy AI for sustainable development, and other international efforts*. President Biden, moreover, used his last address before the UN General Assembly to urge world leaders to "ensure that AI supports, rather than undermines, the core principles that human life has value and all humans deserve dignity."[110]

At the same time that the Biden-Harris Administration was articulating the public rights and values at the center of its AI governance approach, it was also building the science base and practices necessary for agencies and others to test and evaluate risks associated with AI models and systems. Notably, in January 2023, the National Institute of Standards and Technology (NIST) published the Artificial Intelligence Risk Management Framework (AI RMF).[111] The AI RMF's goal is to "offer a resource to the organizations designing, developing, deploying, or using AI systems to help manage the many risks of AI and promote trustworthy and responsible development and use of AI systems."[112] Through the AI RMF, the Administration set out processes that could assist organizations in proactively and systemically identifying and limiting risks—to human rights, safety, as well as the climate—posed by AI systems. It is "voluntary, rights-preserving, non-sector-specific,

109. *Id.*

110. Ja'han Jones, *Biden Warns Dictators Could Use AI to Put 'Shackles' on the 'Human Spirit'*, MSNBC (Sep. 25, 2024), https://www.msnbc.com/the-reidout/reidout-blog/biden-un-speech-ai-rcna172702.

111. NAT'L INST. OF STANDARDS & TECH., U.S. DEP'T OF COM., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0) (Jan. 2023), https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf [hereinafter AI RMF]. This was pursuant to the National Artificial Intelligence Initiative Act of 2020, authorized in division E of the William M. (Mac) Thornberry National Defense Authorization Act for Fiscal Year 2021, Pub. L. No. 116–283, 134 Stat. 3388, which passed with a Congressional override of President Trump's veto.

112. AI RMF, *supra* note 111, at 2.

and use-case agnostic,"[113] and therefore useful to a broad swath of companies, organizations, and federal agencies operating across various sectors of society. It does not disturb existing legal obligations—whether under civil rights, consumer protection, or environmental law—but rather provides process guidance to aid entities attempting to proactively build and deploy systems that reduce the possibility of interfering with rights or obligations, or other objectives an entity might independently seek to achieve.[114] The AI RMF "core" described below also provides useful insights for regulatory and enforcement agencies of the practices for building and deploying AI, identified through an inclusive, multistakeholder process. Such non-binding process standards can inform agencies' regulations, enforcement actions, and remedies.

The AI RMF is divided into two parts: Part I sets the table, describing NIST's view of the risks associated with AI, defining the intended audience for the Framework, and outlining the "characteristics of trustworthy AI systems";[115] and Part II, the "core" of the AI RMF, describes four specific functions (namely, "GOVERN, MAP, MEASURE, AND MANAGE") that the AI RMF's audience can use to address AI risks and improve AI safety and trustworthiness.[116] Notably, Part I defines risk as "the composite measure of an event's probability of occurring and the magnitude or degree of the consequences of the corresponding event,"[117] risk management as "coordinated activities to direct and control an organization with regard to risk,"[118] and posits that the AI RMF offers risk management approaches that both "minimize anticipated negative impacts of AI systems *and* identify opportunities to maximize positive impacts."[119] It further identifies specific

113. *Id.*

114. While some have questioned the use of risk regulation techniques to address AI's impact on rights, that critique is premised on the belief that risks regulation is displacing rights regulation. At least with respect to the AI RMF and recent OMB guidance, this is not the case. The AI RMF is voluntary and non-binding on the private sector, and while OMB MEMO. M-24-10, *supra* note 101, at 16, encouraged agencies to incorporate "additional best practices . . . from the National Institute of Standards and Technology (NIST) AI Risk Management Framework" as well as other sources including the AI BOR, and the revamped OMB guidance on the subject, OMB M-Memo 25-21 incorporates key aspects of it, it does not—and cannot—disturb existing law. The encouragement to adopt practices to systemically and proactively address risks to rights is a complement to U.S. protections for rights not a substitute framework. *See* Margot E. Kaminski, *Regulating the Risks of AI*, 103 B.U. L. REV. 1347, 1378 (2023) (questioning the utility of risk regulation to "*address big, often-unquantifiable, often-contested, often-contextual, and often individualized 'risks'*").

115. AI RMF, *supra* note 111, at 2.

116. *Id.* at 3.

117. *Id.* at 4.

118. *Id.*

119. *Id.*

harms to people, organizations, and ecosystems that AI risk management can prevent.[120] Importantly, the AI RMF "does not prescribe risk tolerance" nor which rights or values to prioritize—rather, it places the onus on implementing organizations to make these determinations, based on contextual and case-or sector-specific factors—informed by regulations, the needs and ethical obligations of relevant professionals, and the needs of affected communities—that can change over time.[121]

Pursuant to EO 14110, described, *supra*, NIST took several additional actions to assist agencies and other organizations in mitigating the risks associated with AI.[122] President Biden established the AI Safety Institute within NIST, which was charged with advancing the science of AI safety and operationalizing capabilities testing on foundational AI models.[123] Other NIST actions included issuing "Guidelines for Evaluating Differential Privacy Guarantees."[124] NIST defines differential privacy as "a privacy-enhancing technology that quantifies privacy risk to individuals when their information appears in a dataset."[125] The guidelines were intended to "help agencies and practitioners of all backgrounds . . better understand how to evaluate promises made (and not made) when deploying differential privacy, including for privacy-preserving machine learning."[126] The guidelines were published alongside a supplemental interactive software archive that "illustrate[s] how to achieve differential privacy and other concepts described in the publication."[127]

120. *Id.* at 5.

121. *Id.* at 7.

122. Exec. Order No. 14,110 §§ 4.1, 4.4.b.ii, 9.b., 10.1.d.i, 88 Fed. Reg. 75191, 75196, 75201, 75217, 75219 (Nov. 1, 2023) (signed Oct. 30, 2023).

123. *FACT SHEET: Vice President Harris Announces New U.S. Initiatives to Advance the Safe and Responsible Use of Artificial Intelligence*, WHITE HOUSE (Nov. 1, 2023), https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2023/11/01/fact-sheet-vice-president-harris-announces-new-u-s-initiatives-to-advance-the-safe-and-responsible-use-of-artificial-intelligence/ (reporting that the Vice President is announcing the creation of The United States AI Safety Institute (US AISI) inside NIST). For an overview of AISI, see U.S. A.I. SAFETY INST., NAT'L INST. OF STANDARDS & TECH., THE UNITED STATES ARTIFICIAL INTELLIGENCE SAFETY INSTITUTE: VISION, MISSION, AND STRATEGIC GOALS (May 21, 2024), https://www.nist.gov/system/files/documents/2024/05/21/AISI-vision-21May2024.pdf.

124. Joseph Near & David Darais, *NIST SP 800-226 (Initial Public Draft) Guidelines for Evaluating Differential Privacy Guarantees*, NAT'L INST. OF STANDARDS & TECH. (Dec. 11, 2023), https://csrc.nist.gov/pubs/sp/800/226/ipd.

125. *Id.* at 1.

126. *Id.*

127. *Id.* (describing Python Jupiter Notebook packages that illustrate how to achieve differential privacy and other concepts described in the publication); *see also* NAT'L INST. OF STANDARDS & TECH., U.S. DEP'T OF COM., *NIST-SP-800-226-SupplementalMaterial*, GITHUB (Dec. 11, 2023), https://github.com/usnistgov/PrivacyEngCollabSpace/tree/master/tools/de-identification/NIST-SP-800-226-SupplementalMaterial/.

NIST also published the "Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile," a companion resource to the AI RMF.[128] This profile is an implementation of the Framework's functions, categories, and subcategories in the context of GAI, and intends to assist organizations in "deciding how to best manage AI risks in a manner that is well-aligned with their goals, consider[ing] legal/regulatory requirements and best practices, and reflect[ing] risk management priorities."[129] It describes the time scale, scope, and potential sources of risks from GAI, but—like the Framework itself—leaves it to organizations to determine how best to measure and tolerate GAI risks.[130] And in November 2024, NIST released a report on "existing standards, tools, methods, and practices"[131] for:

> authenticating [synthetic] content and tracking its provenance; labeling synthetic content, such as using watermarking; detecting synthetic content; preventing generative AI (GAI) from producing child sexual abuse material or producing non-consensual intimate imagery of real individuals (to include intimate digital depictions of the body or body parts of an identifiable individual); testing software used for the above purposes; and auditing and maintaining synthetic content.[132]

The report outlines potential risks and harms associated with synthetic content, including "the target audience for the content; the context in which content is used or misused; the sophistication of the actor creating and/or disseminating the content; and any social, economic, and health-related (including mental health) costs" associated with its dissemination.[133]

The Biden-Harris Administration complemented this non-exhaustive list of guidance and tooling targeted toward GAI risk management with regulations aimed at both protecting rights and mitigating risks in particular sectors. One particularly illustrative example is the Department of Health and Human Services' (HHS) July 2024 rule on bias in AI systems.[134] Promulgated

128. NAT'L INST. OF STANDARDS & TECH., U.S. DEP'T OF COM., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK: GENERATIVE ARTIFICIAL INTELLIGENCE PROFILE 1 (July 2024), https://doi.org/10.6028/NIST.AI.600-1.

129. *Id.* at 1.

130. *Id.* (explaining that the profile is designed to assist organizations in managing AI risks in a manner that is aligned with the organization's risk tolerance, and resources).

131. NAT'L INST. STANDARDS & TECH., U.S. DEP'T OF COM., REDUCING RISKS POSED BY SYNTHETIC CONTENT AN OVERVIEW OF TECHNICAL APPROACHES TO DIGITAL CONTENT TRANSPARENCY 1 (Nov. 20, 2024), https://www.nist.gov/publications/reducing-risks-posed-synthetic-content-overview-technical-approaches-digital-content.

132. *Id.*

133. *Id.* at 2.

134. Nondiscrimination in Health Programs and Activities, 89 Fed. Reg. 37522 (May 6, 2024) (codified as amended at 42 U.S.C. § 18116).

pursuant to Section 1557 of the Affordable Care Act, this regulation aimed to combat bias in the AI systems used by healthcare providers and insurers for clinical care and health administration.[135] Section 1557 of the Affordable Care Act prohibits bias on the basis of "race, color, national origin, sex, age, or disability,"[136] and the new rule requires that covered entities take "reasonable steps" to "identify and mitigate potential discrimination" from any "augmented decision-making tools or models, such as artificial intelligence (AI) and machine learning" deployed in a health care or health insurance setting.[137] If covered entities do not take such "reasonable steps," HHS may take "corrective action" against them.[138]

This HHS rule exemplifies the Biden-Harris Administration's layered, complementary approach to AI regulation. The Administration, in the many executive orders, memos, and public statements referenced, *supra*, made clear that equity, human rights, and democratic values were at the core of its approach to AI governance. The Administration highlighted sectors like healthcare and education where laws offered protection, but regulations needed updating to ensure that the public's rights and safety would be protected. It laid out frameworks and practices for mitigating risks associated with AI—including privacy, generative AI, and synthetic content risks that would affect multiple sectors—but maintained the regulatory and enforcement authority of existing expert agencies. The frameworks and tools help regulated and unregulated entities mitigate risks, but do not prescriptively dictate their activities. It directed expert federal agencies to issue specific regulations that spell the ways to address AI risks to important rights at the sector level, such as the right to nondiscrimination in healthcare, and required regulated entities to mitigate risks to those rights.

In essence, the Biden-Harris Administration's regulatory approach to AI was neither "risk-based" or "rights-based," but incorporated principles from both approaches. It fostered the development of specialized frameworks and methods for evaluating AI risks, but because the rights at stake vary across sectors, the Administration empowered sector-specific agencies to establish how those approaches should be used to address AI-specific risks, such as the right to nondiscrimination in healthcare. Maintaining the emphasis on the substantive domains of use centers rights-expertise and positions technical experts and knowledge as a facilitator of the government's field-specific agencies responsible for protecting rights and safety. This "field-centric" approach to AI governance maintains the centrality of the agencies' missions,

135. *Id.* at 37522.
136. *Id.*
137. *Id.* at 37524, 37642.
138. *Id.* at 37524, 37557.

rather than emphasizing new technology entering the regulated market. It avoids technosolutionism—constructing complex phenomena as *problems* that technology is best able to solve.[139] And it corrects the misguided assumption that the use of AI reduces the relevance of existing regulations. Just as companies' use of data analytics, data mining, and statistical models in regulated areas must comply with the law, so too must the use of AI.

B. BRINGING EXPERTISE AND CAPACITY INTO GOVERNMENT: DIRECT HIRING AND STAKEHOLDER INVOLVEMENT

The Administration used the primacy of existing domain specific regulatory and enforcement agencies and addressed the need for specialized expertise. As this section describes, EO 14110 took several steps to ensure federal agencies had the AI and AI-enabling talent to design and use AI, and to effectively regulate its use. This included upskilling staff across agencies, bringing new technical professionals into key service delivery and enforcement agencies, and creating the AI Safety Institute within the National Institutes of Standards and Technology with new AI specific staff. In addition, the Administration helped scaffold the participation of a broad range of stakeholders in AI governance activities and broadened financial investments in the field of public interest technology to bolster the expertise in civil society organizations, train future generations of public interest and mission-oriented technologists, and support research to protect the public's rights and safety.

1. Bringing AI and AI Enabling Talent into Federal Service

The AI and Tech Talent Task Force, created by the President through EO 14110, launched an AI Talent Surge to accelerate hiring AI and AI-enabling professionals across the federal government.[140] This effort included: flexible hiring authorities for federal agencies to bring in AI talent, including direct hire authorities and excepted service authorities;[141] an interagency working group

139. Technosolutionism is both the mindset or belief that complex societal problems can be solved by technology and the construction of phenomena or things as problems which technology is best situated to solve. EVGENY MOROZOV, TO SAVE EVERYTHING, CLICK HERE: THE FOLLY OF TECHNOLOGICAL SOLUTIONISM 6 (2013) ("Solutionism . . . is not just a fancy way of saying that for someone with a hammer, everything looks like a nail; it's not just another riff on the inapplicability of 'technological fixes' to 'wicked problems' . . . It's not only that many problems are not suited to the quick-and-easy solutionist tool kit. It's also that what many solutionists presume to be 'problems' in need of solving are not problems at all.").

140. AI & TECH TALENT TASK FORCE, INCREASING AI CAPACITY ACROSS THE FEDERAL GOVERNMENT: AI TALENT SURGE PROGRESS AND RECOMMENDATIONS (Apr. 26, 2024), https://digital.library.unt.edu/ark:/67531/metadc2349490/m2/1/high_res_d/AI-Talent-Surge-Progress-Report.pdf.

141. Memorandum from Kiran A. Ahuja, Dir., U.S. Off. Personnel Mgmt., on Government-Wide Hiring Authorities for Advancing Federal Government Use of Artificial Intelligence (AI) to Heads of Departments and Agencies 1 (Dec. 29, 2023), https://

of human resources professionals, recruiting experts, technical leads, and hiring managers to share best practices on federal government-wide hiring of people with AI and other technical skills;[142] guidance to agencies on how to assess and expand the AI competencies of their staff through recruitment, selection, and hiring;[143] and guidance on pay flexibility, incentive pay, and leave and workforce flexibility programs.[144] It also scaled up the use of government-wide tech talent programs, including the Presidential Innovation Fellows, U.S. Digital Corps, and U.S. Digital Service and the creation of the new DHS AI Corps.[145] Through these programs, the Administration made over 200 hires by the end of July 2024.[146] The White House AI and Tech Talent Task Force coordinated the processes and oversaw the distribution of many of the new professionals across federal agencies.

The Administration explicitly provided guidance that defined the skills and competencies required in the AI workforce. Importantly, that guidance clarified that key technical competencies included "[s]ociotechnical [s]ystems," "[t]esting and [v]alidation," and "[v]alues-[d]riven [d]esign."[147] It defined "values-driven design" as:

> [S]ystematically applies principles and techniques from relevant subject matter domains to all aspects of design, development, maintenance, and deployment to protect the rights and safety of stakeholders and the public, ensuring equity, security, privacy, autonomy, accessibility, justice, beneficence, and nonmaleficence. Creatively combines technical and policy approaches to protect and support these core values. [And e]nsures that values inform the design, deployment, testing, and oversight of AI systems, and that

www.opm.gov/chcoc/transmittals/2023/Government-wide%20Hiring%20Authorities%20for%20Advancing%20Federal%20Government%20Use%20of%20Artificial%20Intelligence%20(AI)%2012-29-2023.pdf.

142. AI & TECH TALENT FORCE, *supra* note 140, at 6.

143. Ahuja, *supra* note 141, at 1.

144. Memorandum from Kiran A. Ahuja, Dir., U.S. Off. Personnel Mgmt., on Skills-Based Hiring Guidance and Competency Model for Artificial Intelligence Work to Heads of Executive Departments and Agencies (Apr. 29, 2024), https://www.opm.gov/chcoc/transmittals/2024/Skills-Based%20Hiring%20Guidance%20and%20Competency%20Model%20for%20Artificial%20Intelligence%20Work.pdf.

145. AI & TECH TALENT FORCE, *supra* note 140, at 3, 6.

146. *FACT SHEET: Biden-Harris Administration Announces New AI Actions and Receives Additional Major Voluntary Commitment on AI*, WHITE HOUSE 7 (July 26, 2024), https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2024/07/26/fact-sheet-biden-harris-administration-announces-new-ai-actions-and-receives-additional-major-voluntary-commitment-on-ai/.

147. Ahuja, *supra* note 144, at 3.

> important value-related design choices are communicated to end users.[148]

This standard signaled an expectation that the technical workforce enlisted to build the government's capacity to use and regulate AI would have an understanding of the politics of technical artifacts[149] and how to construct and use them to protect core rights and values,[150] in addition to more standard competencies such as data visualization and machine learning. This guidance assisted agencies in identifying key skills and competencies needed for AI professionals, increased opportunities for individuals with nontraditional academic backgrounds,[151] and broadened agencies' and the public's understanding of the skills necessary for responsible development of AI.

The Task Force worked to bolster the technical expertise at agencies attempting important AI-use cases and those with significant enforcement missions and short on technical expertise. Some agencies had funding to hire additional staff using direct hire authorities. Most notably, the Department of Homeland Security brought on nearly fifty new AI and AI-enabling personnel to work on missions, such as countering fentanyl networks, combating child sexual exploitation and abuse, and delivering immigration services.[152] In addition, the Administration launched a novel effort to build a pool of Science, Technology, Engineering, and Math (STEM) and AI experts that could be flexibly deployed to help agencies support implementation of the EO 14110, as well as the National Security Memorandum on Revitalizing America's

148. *Id.* at 16.

149. Winner, *supra* note 60, at 121–36.

150. Katie Shilton, Jes A. Koepfler & Kenneth R. Fleischmann, *How to See Values in Social Computing: Methods for Studying Values Dimensions*, CSCW '14: PROCS. OF THE 17TH ACM COMPUT. SUPPORTED COOP. WORK & SOC. COMPUTING, 426–35 (2014); Cory Knobel & Geoffrey C. Bowker, *Values in Design*, 54 COMMC'NS ACM 26 (2011); Mary Flanagan, Daniel C. Howe & Helen Nissenbaum, *Embodying Values in Technology: Theory and Practice*, *in* INFORMATION TECHNOLOGY AND MORAL PHILOSOPHY 322, 322 (Jeroen van den Hoven & John Weckert eds., 2008); Deirdre K. Mulligan & Helen Nissenbaum, *The Concept of Handoff as a Model for Ethical Analysis and Design, in* OXFORD HANDBOOK OF ETHICS OF AI 232, 233 (Markus D. Dubber, Frank Pasquale & Sunit Das eds., 2020); *see generally* MARY FLANAGAN & HELEN NISSENBAUM, VALUES AT PLAY IN DIGITAL GAMES (2014) (developing a framework for identifying socially recognized moral and political values in technology in the context of digital games).

151. Ahuja, *supra* note 144, at 1 ("OPM is pleased to issue skills-based hiring guidance and a competency model for Artificial Intelligence (AI), data, and technology talent to assist agencies to identify key skills and competencies needed for AI professionals and increase access to these technical roles for individuals with nontraditional academic backgrounds.").

152. *DHS Generative AI Sector Playbook*, DEP'T OF HOMELAND SEC., https://www.dhs.gov/ai (last visited Aug. 21, 2024); Justin Doubleday, *DHS Sets 'Aggressive' Recruiting Strategy to Fill AI Jobs*, FED. NEWS NETWORK (Feb. 19, 2024), https://federalnewsnetwork.com/artificial-intelligence/2024/02/dhs-sets-aggressive-recruiting-strategy-to-fill-ai-jobs/.

Foreign Policy and National Security Workforce, Institutions, and Partnerships (NSM-3) and other presidential priorities.[153] The Department of Defense, with support from the OMB and the OSTP announced a new program which, if launched, would provide a reserve team of STEM and AI experts from academia and elsewhere that agencies could tap for short-term engagements to bring appropriate expertise in to assist with implementations, evaluations, and other work.[154] Centralizing the work of hiring and clearing these advisors and providing a simple mechanism for a range of agencies to bring them in for specific projects would expand the range of experts agencies could practically and financially afford to bring into their efforts.

Enforcement agencies with more dedicated technical expertise produced reports and held workshops to assist peer agencies; led calls to action clarifying the importance and need for technical experts; and created networks to build momentum domestically and internationally for increased technical expertise within consumer protection, competition, and civil rights enforcement agencies. For example, the FTC's Office of Technology issued a report designed to "establish a shared context and serve as a resource for building technical capacity in government agencies" and share information about how the Office of Technology "applies subject matter experts in regulatory and enforcement contexts."[155] The FTC took this effort to the international stage, initiating the International Competition Network Tech Forum's work to define best practices in building tech capacity in law enforcement agencies.[156] The Consumer Financial Protection Bureau's Office of Technology similarly sought to boost technical expertise across enforcement agencies.[157] They

153. *Fact Sheet: Biden-Harris Administration Announces Commitments from Across Technology Ecosystem Including Nearly $100 Million to Advance Public Interest Technology*, WHITE HOUSE (July 16, 2024), https://bidenwhitehouse.archives.gov/ostp/news-updates/2024/07/16/fact-sheet-biden-harris-administration-announces-commitments-from-across-technology-ecosystem-including-nearly-100-million-to-advance-public-interest-technology/.

154. *Id.* (announcing the Trusted Advisors Pilot). This program was not operational by the end of the Biden-Harris Administration, and it is unclear whether work to stand it up is continuing or if the Trump Administration has abandoned it.

155. OFF. TECH. STAFF, FED. TRADE COMM'N, BUILDING TECH CAPACITY IN LAW ENFORCEMENT AGENCIES 3 (Mar. 2024), https://www.ftc.gov/system/files/ftc_gov/pdf/ot.techcapacityreport.pdf.

156. OFF. OF TECH., *Best Practices in Building Tech Capacity in Law Enforcement Agencies*, FED. TRADE COMM'N (Mar. 26, 2024), https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2024/03/best-practices-building-tech-capacity-law-enforcement-agencies; *see also Building Digital Capacity to Strengthen and Support Law Enforcement Agencies*, INT'L COMPETITION NETWORK, https://www.internationalcompetitionnetwork.org/working-groups/icn-operations/technologists/technologist-forum-statement-on-building-agency-digital-capacity/ (last visited Sep. 19, 2025).

157. Erie Meyer, *Public Interest Tech Jobs: Regulate Tech and AI*, CONSUMER FIN. PROT. BUREAU (May 20, 2024), https://www.consumerfinance.gov/about-us/blog/public-interest-tech-jobs-regulate-tech-and-ai/ (announcing the launch of a cross-government hiring effort

hosted trainings and briefings for enforcement agencies on numerous consumer financial protection topics, including emerging practices in biometrics and how to address algorithmic harms[158] and authored a guide to help enforcement agencies hire technologists to protect consumers.[159] In addition, EO 14110 directed the Civil Rights Division of the Department of Justice to convene federal civil rights offices to advance comprehensive use of their respective authorities and offices to prevent and address discrimination in the use of automated systems, including algorithmic discrimination, and "develop, as appropriate, additional training, technical assistance, guidance, or other resources . . . and consider providing [similar support] to State, local, Tribal, and territorial investigators and prosecutors."[160]

The White House and federal agencies took many other actions to build the staff of sociotechnical experts within agencies and directly available to them to support the responsible use and governance of AI.

2. Building the Responsible AI Field

The Administration further sought to build the field of responsible AI generally, rather than just the government capacity. The Administration established the U.S. AI Safety Institute (AISI), which is housed within NIST, to advance the science of AI safety; articulate, demonstrate, and disseminate the practices of AI safety; and support institutions, communities, and coordination around AI safety.[161] To achieve these goals, the AISI conducts testing of advanced models and systems to assess potential and emerging risks; develops guidelines on evaluations and risk mitigations; and performs and

to embed technical experts across multiple agencies that share a variety of consumer protection, competition, and civil rights authorities).

158. Erie Meyer, *Bringing Tech Enforcers Together to Protect Consumers*, CONSUMER FIN. PROT. BUREAU (Mar. 14, 2023), https://www.consumerfinance.gov/about-us/blog/bringing-tech-enforcers-together-to-protect-consumers/.

159. *Hiring Technologists to Protect Consumers*, CONSUMER FIN. PROT. BUREAU, https://www.consumerfinance.gov/about-us/careers/cfpb-technologist/hiring-technologists-to-protect-consumers/ (last modified Oct. 30, 2024).

160. Exec. Order No. 14,110 § 7.1(ii)–(iii), 88 Fed. Reg. 75191, 75211 (Nov. 1, 2023) (signed Oct. 30, 2023).

161. Press Release, Off. of Pub. Affs., at the Direction of President Biden, Department of Commerce to Establish U.S. Artificial Intelligence Safety Institute to Lead Efforts on AI Safety (Nov. 1, 2023), https://www.commerce.gov/news/press-releases/2023/11/direction-president-biden-department-commerce-establish-us-artificial. The Trump Administration renamed it to the Center for AI Standards and Innovation and narrowed its agenda. Press Release, Off. of Pub. Affs., Statement from U.S. Secretary of Commerce Howard Lutnick on Transforming the U.S. AI Safety Institute into the Pro-Innovation, Pro-Science U.S. Center for AI Standards and Innovation (June 3, 2025), https://www.commerce.gov/news/press-releases/2025/06/statement-us-secretary-commerce-howard-lutnick-transforming-us-ai.

coordinates technical research.[162] To enable this work, AISI works closely with experts from across the AI industry, civil society, and sister safety institutes.[163]

AISI and the broader set of NIST experts working on AI were tasked with providing technical guidance to support the responsible development and use of AI. As noted above, AISI and NIST issued guidance—some still in progress—on a range of technical issues, including guidance for AI developers in managing the evaluation of misuse of dual-use foundation models, frameworks on managing generative AI risks and securely developing generative AI systems and dual-use foundation models, and provided a technical report to the White House outlining tools and techniques to reduce the risks from synthetic content.[164] All of these documents were produced with input from stakeholders.

To support stakeholder participation in AI governance, AISI established a consortium of over 200 AI stakeholders that seeks to "unite AI creators and users, academics, government and industry researchers, and civil society organizations in support of the development and deployment of safe and trustworthy artificial intelligence."[165] The consortium includes a wide range of stakeholders including AI companies like Anthropic and OpenAI, technology companies like Apple and Google, energy companies including PG&E, chip manufacturers like NVIDIA, and universities including NYU, Syracuse, and UC Berkeley.[166] The consortium has working groups on topics including Risk Management for Generative AI, Synthetic Content, Capability Evaluations, Red-Teaming, and Safety & Security.[167]

AISI began to address the access challenges that stymie stakeholders' full participation in decisions about AI models. AISI signed memorandums of understanding (MOU) with two major AI companies, Anthropic and OpenAI, that "enable formal collaboration on AI safety research, testing and

162. *See generally* U.S. A.I. SAFETY INST., NAT'L INST. OF STANDARDS & TECH., *supra* note 123.

163. *Id.* at 2.

164. *See, e.g.*, *Department of Commerce Announces New Guidance, Tools 270 Days Following President Biden's Executive Order on AI*, NAT'L INST. OF STANDARDS & TECH. (July 26, 2024), https://www.nist.gov/news-events/news/2024/07/department-commerce-announces-new-guidance-tools-270-days-following.

165. *Biden-Harris Administration Announces First-Ever Consortium Dedicated to AI Safety*, NAT'L INST. OF STANDARDS & TECH. (Feb. 8, 2024), https://www.nist.gov/news-events/news/2024/02/biden-harris-administration-announces-first-ever-consortium-dedicated-ai.

166. *Artificial Intelligence Safety Institute Consortium: AISIC Members*, NAT'L INST. OF STANDARDS & TECH., https://www.nist.gov/artificial-intelligence/artificial-intelligence-safety-institute-consortium-aisic/aisic-members (last visited Sep. 17, 2025).

167. *Artificial Intelligence Safety Institute Consortium: AISIC Working Groups*, NAT'L INST. OF STANDARDS & TECH., https://www.nist.gov/artificial-intelligence/artificial-intelligence-safety-institute-consortium-aisic/aisic-working (last visited Sep. 17, 2025).

evaluation" between these companies and the federal government.[168] Each company's MOU "establishes the framework for the U.S. AI Safety Institute to receive access to major new models from each company prior to and following their public release" to enable collaborative research on AI model capabilities, safety risks, and risk mitigation.[169]

To advance global coordination on AI governance, AISI hosted the inaugural convening of an International Network of AI Safety Institutes in San Francisco. Secretary of Commerce Gina Raimondo welcomed delegations from AI Safety Institutes in Australia, Canada, the European Union, France, Japan, Kenya, the Republic of Korea, Singapore, and the United Kingdom.[170] The goal of the convening was to "kickstart the Network's technical collaboration ahead of the AI Action Summit in Paris in February 2025."[171] The convening also included "experts from international civil society, academia, and industry" who would "inform the work of the Network and ensure a robust view of the latest developments in the field of AI."[172]

In the leadup to the convening, AISI formed the Testing Risks of AI for National Security (TRAINS) Taskforce, which is chaired by AISI and includes representation from the Department of Defense, including the Chief Digital and Artificial Intelligence Office and the National Security Agency; the Department of Energy and ten of its National Laboratories; the Department of Homeland Security, including the Cybersecurity and Infrastructure Security Agency; and the National Institutes of Health at the Department of Health and Human Services.[173] TRAINS aimed to enable coordinated research and testing of advanced AI models on issues related to national security, including "radiological and nuclear security, chemical and biological security, cybersecurity, critical infrastructure, [and] conventional military capabilities."[174]

168. *U.S. AI Safety Institute Signs Agreements Regarding AI Safety Research, Testing and Evaluation with Anthropic and OpenAI*, NAT'L INST. OF STANDARDS & TECH. (Aug. 29, 2024), https://www.nist.gov/news-events/news/2024/08/us-ai-safety-institute-signs-agreements-regarding-ai-safety-research.

169. *Id.*

170. Press Release, U.S. Dep't of Com., U.S. Secretary of Commerce Raimondo and U.S. Secretary of State Blinken Announce Inaugural Convening of International Network of AI Safety Institutes in San Francisco (Sep. 18, 2024), https://www.commerce.gov/news/press-releases/2024/09/us-secretary-commerce-raimondo-and-us-secretary-state-blinken-announce.

171. *Id.*

172. *Id.*

173. Press Release, U.S. Dep't of Com., U.S. AI Safety Institute Establishes New U.S. Government Taskforce to Collaborate on Research and Testing of AI Models to Manage National Security Capabilities & Risks (Nov. 20, 2024), https://www.commerce.gov/news/press-releases/2024/11/us-ai-safety-institute-establishes-new-us-government-taskforce.

174. *Id.*

The Biden-Harris Administration invested in building the AI research ecosystem necessary to support the use of AI for important public missions and to establish a strong, sociotechnical understanding of risks and the methods and tools to address them. It created eighteen new AI institutes across the United States through the National Science Foundation-led National Artificial Intelligence Research Institutes program—a research investment that began in August of 2020 during the Trump-Pence Administration which established the first seven national AI research institutes.[175] Many of these institutes enjoy private support as well, but the public investment ensures they are directed towards research that will benefit the public, ranging from promoting ethical and trustworthy AI systems and technologies, developing novel approaches to cybersecurity, addressing climate change, expanding our understanding of the brain, and enhancing education and public health.[176] In addition, it launched the National Science Foundation's (NSF) Responsible Design, Development, and Deployment of Technologies (ReDDDoT) program that supports multidisciplinary, multi-sector teams that examine and demonstrate the principles, methodologies, implementations, and impacts associated with responsible design, development, and deployment of technologies in practice.[177] A collaboration between the NSF and philanthropic funders, the program and other publicly funded efforts support research that is developing new methods and approaches to ensure that ethical, legal, and societal considerations and community values are embedded across technology lifecycles to generate products that promote the public's well-being and mitigate harm.

The Administration launched the National AI Research Resource (NAIRR) pilot—a national infrastructure led by the NSF in partnership with the Department of Energy and other governmental and nongovernmental partners—that makes available resources to support the nation's AI research and education community.[178] The NAIRR supports research teams across forty-nine states that are tackling projects covering deepfake detection, AI safety, next-generation medical diagnoses, environmental protection, and

175. Michael Kratsios & Chris Liddell, Off. Sci. & Tech. Pol'y, *The Trump Administration Is Investing $1 Billion in Research Institutes to Advance Industries of the Future*, WHITE HOUSE (Aug. 26, 2020), https://trumpwhitehouse.archives.gov/articles/trump-administration-investing-1-billion-research-institutes-advance-industries-future/.

176. *NSF Announces $100 Million Investment in National Artificial Intelligence Research Institutes Awards to Secure American Leadership in AI*, NAT'L SCI. FOUND. (July 29, 2025), https://www.nsf.gov/news/nsf-announces-100-million-investment-national-artificial (discussing public-private partnerships).

177. *Responsible Design, Development, and Deployment of Technologies*, NAT'L SCI. FOUND. (Jan. 8, 2024), https://www.nsf.gov/funding/opportunities/redddot-responsible-design-development-deployment-technologies/506215/nsf24-524.

178. NAT'L A.I. RSCH. RES. PILOT, https://nairrpilot.org (last visited Sep. 17, 2025).

materials engineering.[179] It provides public infrastructure to support research necessary to address AI governance challenges.[180]

These public research investments are essential for progress on AI Governance. For example, the Defense Advanced Research Projects Agency (DARPA) launched one of the first and most recognized programs in the area of Explainable AI (XAI) with the goal of enabling end users to better understand, trust, and effectively manage artificially intelligent systems.[181] This early public research investment sparked broad interest in an area essential to AI governance.[182] And this is just one example. A robust publicly funded research ecosystem will produce the methods and tools to ensure AI systems are fit for purpose, trustworthy, rights-respecting, and safe. It will also ensure AI research advances our nation's grand ambitions and addresses our gravest risks, whether they arise from adversarial nations seeking to undermine our national security, or decades of inaction to address the looming climate crisis.

Together these actions bolstered both the technical workforce and technical expertise in government, as well as in civil society, academia, and other stakeholders. The Administration invested in the research and education to create AI governance methods and future practitioners, galvanized support for the field of public interest technology, provided resources to support AI research and applications outside large companies, and created momentum for similar international efforts. The Administration created public venues for all stakeholders to participate in "governance-by-design" work. These efforts help ensure AI governance is consistent with the norms of public governance and designed to center and be responsive to public values and not just private interests.

---

179. To see the range of projects supported, see *Resource Allocation*, NAT'L A.I. RSCH. RES. PILOT, https://nairrpilot.org/projects/awarded (last visited Sep. 17, 2025).

180. *Democratizing the Future of AI R&D: NSF to Launch National AI Research Resource Pilot*, NAT'L SCI. FOUND. (Jan. 24, 2024), https://www.nsf.gov/news/democratizing-future-ai-rd-nsf-launch-national-ai-research.

181. *XAI: Explainable Artificial Intelligence*, DEF. ADVANCED RSCH. PROJECTS AGENCY, https://www.darpa.mil/research/programs/explainable-artificial-intelligence (last visited Sep. 17, 2025).

182. David Gunning, Eric Vorm, Jennifer Yunyan Wang & Matt Turek, *DARPA's Explainable AI (XAI) Program: A Retrospective*, 2 APPLIED AI LETTERS e61 (2021) (claiming the program stimulated the field of explainable AI research and "produced a more nuanced understanding of XAI uses and users, the psychology of XAI, the challenges of measuring explanation effectiveness, as well as producing a new portfolio of XAI ML and HCI techniques"); Atul Rawal, James McCoy, Danda B. Rawat, Brian M. Sadler & Robert St. Amant, *Recent Advances in Trustworthy Explainable Artificial Intelligence: Status, Challenges, and Perspectives*, 3 IEEE TRANSACTIONS ON A.I. 852, 852–53 (2022) (Figure 1 showing the rise in explainable AI research papers in 2017 and noting its "emergence along with the U.S. DoD DARPA XAI program").

C. MAINTAINING THE PUBLICNESS OF POLICYMAKING: FOCUSING ON IMPACT RATHER THAN SYSTEMS AND REQUIRING STAKEHOLDER PARTICIPATION THROUGHOUT THE AI LIFECYCLE

The Administration sought to maintain the publicness of the values and policies embedded in AI systems and provide robust opportunities for public input into its deliberations to intervene in technological design. It did so in general and specific ways.

This Section first discusses two interventions that reframed the project of AI governance away from models and systems towards impacts on individuals' and communities' rights and safety: the Blueprint for an AI Bill of Rights[183] and the turn to *use cases* as the focus of evaluation in federal AI governance. This Section next describes in detail how this reframing along with new system documentation and public engagement requirements in the OMB guidance to federal agencies on the development and use of AI maintained the visibility and attention to rights and values during agency design and deployment practices. Finally, this Section documents the Administration's approach to a significant policy—the availability of model-weights for powerful AI models—which engaged the public in considerations of the breadth of values implicated in a governance-by-design strategy and produced a policy that exhibits modesty and restraint in design.

1. Reframing The Project of AI Governance

Two overarching and underappreciated interventions reoriented the debates about AI governance: the Blueprint for the AI Bill of Rights[184] and the use-case orientation of the federal guidance on the development and use of AI.

a) The AI Bill of Rights

First, as Alondra Nelson, former Principal Deputy Director for Science and Society at the White House OSTP, Deputy Assistant to the President, who led the creation of the White House Blueprint for an AI Bill of Rights (AI BoR), explained, the AI BoR acts as "civic architecture" "creating infrastructure for collective participation in AI policy."[185] It grounded AI governance in the achievement of rights and civil liberties, establishing "a sociotechnical approach that recalibrates the relationship between technology

183. AI BOR, *supra* note 87.

184. *Id.*

185. Alondra Nelson, *From Blueprint to Building Blocks: The AI Bill of Rights as Civic Architecture* 2 (forthcoming) (on file with authors); *see also* Alondra Nelson, Inst. for Advanced Study, Presentation at Artificial Intelligence and Democratic Freedoms, Symposium by Knight First Amendment Institute at Columbia University (Apr. 10, 2025), https://knightcolumbia.org/events/artificial-intelligence-and-democratic-freedoms.

and society by positioning individuals not merely as passive users of AI systems but as rights-bearing individuals and communities with legitimate claims to protection, agency, and redress."[186] The AI BoR did more than establish guiding principles, it established a frame and reference point for the public to see themselves, their communities, and the stakes in what were technocratic, expert-dominated debates.

Second, and perhaps stealthier to those outside the government, the OMB guidance to agencies on the responsible development and use of AI focused on the evaluation of AI *use cases*. Typically, the object of assessment or evaluation is a *system*[187] generally exclusive to the technical aspects, not the institutional aspects, that together co-create a system's ultimate impact. By reorienting the government's analysis of AI around *use cases*, such as use of an AI to assist in benefits determinations, the Administration established a new paradigm for accounting for the potential outcomes of incorporating AI into a government process. This reorientation adopts a sociotechnical approach to risk management found in high-risk fields.[188]

Focusing on use cases addresses shortcomings that frequently pervade other methods of assessment. As technology scholar Roel Dobbe has explained, safety science research reveals that "systems cannot be safeguarded by technical design choices on the model or algorithm alone."[189] Rather, he explains, it is necessary to take an "end-to-end" approach to analyzing risks and a sociotechnical system—a perspective that considers "the context of use, impacted stakeholders . . . and informal institutional environment" when deploying mitigations.[190]

186. Nelson, *From Blueprint to Building Blocks*, *supra* note 185, at 1.

187. For example, 44 U.S.C. § 3501 note (2000 & 2002 Amendments), and OFF. OF MGMT. & BUDGET, EXEC. OFF. OF THE PRESIDENT, MEMO. NO. M-03-22, OMB GUIDANCE FOR IMPLEMENTING THE PRIVACY PROVISIONS OF THE E-GOVERNMENT ACT OF 2002 (Sep. 26, 2003) (requiring agencies to conduct a Privacy Impact Assessment before "developing or procuring information technology that collects, maintains, or disseminates information that is in an identifiable form").

188. One notable exception to the use-case orientation is the Administration's focus on the risks posed by dual-use foundation models with widely available model weights, discussed, *infra*, in Part IV(C)(2), which takes a systems analysis. However, as explained, *infra*, the inquiry led by the National Telecommunications and Information Administration (NTIA) explored contextualized risks and benefits of systems in various domains of use, again centering expected impact rather than abstract risk.

189. Roel I. J. Dobbe*, System Safety and Artificial Intelligence*, *in* THE OXFORD HANDBOOK OF AI GOVERNANCE 441*,* 441 (Justin B. Bullock et al., eds., 2022).

190. *Id.*

By considering policies, users, the technical interfaces, and outputs in context, a use-case orientation thus avoids "traps"[191] researchers have identified in detached evaluation of technical systems alone. Specifically, such a sociotechnicalsystem framing helps resist abstractions common in technical practice that push important normative decisions out of view,[192] rendering visible built-in politics and values, and opening them up for contestation.[193]

Such visibility and contestation is necessary to ensure a systemic orientation towards public values, and away from what Cohen and Waldman have called regulatory managerialism—the importing of the "practices for organizing and overseeing private sector, capitalist economic production and . . . the logics and underlying ideologies in which those practices are rooted" into regulated activities.[194] As scholar of regulation Christie Ford has powerfully argued, regulatory managerialism contributes to "peoples' . . . alienation from public institutions, and the perspectives of the regulators who are supposed to be safeguarding their interests."[195]

The focus on *use cases* thus responds to the limits of algorithmic accountability,[196] and calls to introduce more qualitative elements into assessments,[197] including scenario analysis, which produces an "extended narrative prediction of how a given policy decision will increase the likelihood of some complex set of consequences and decrease that of others."[198] Such tools "enable policy evaluators to better understand and predict interrelated aspects of technological advance, economic changes, and policy shifts,"[199]

---

191. Andrew D. Selbst, Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian & Janet Vertesi, *Fairness and Abstraction in Sociotechnical Systems*, FAT* '19: PROCS. OF THE CONF. ON FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY 59 (2019).

192. *Id.*

193. *See* Mulligan & Bamberger, *Procurement as Policy*, *supra* note 1, at 842–57 (discussing the importance of political-visibility-enhancing processes and the resulting contestability).

194. Cohen & Waldman, *supra* note 26, at iv.

195. Christie Ford, *Regulation as Respect*, 86 L. & CONTEMP. PROBS. 133, 134 (2023).

196. *See id.* at 138 ("[Regulators] tend not even to have methods for determining which specific groups of people—especially vulnerable ones, however defined—should be considered as part of any regulatory impact assessment. Nor do they have decision rules for how to determine the balances between benefits and costs when considering disaggregated groups. These gaps highlight the level of abstraction from real humans at which managerialism operates, and they make inequities harder to see.").

197. Andrew D. Selbst, *An Institutional View of Algorithmic Impact Assessments*, 35 HARV. J. L. & TECH. 117, 122, 180 (2021); *see also* Rory Van Loo, *Stress Testing Governance*, 75 VAND. L. REV. 553, 558 (2022) (advocating the use of "stress tests" that "incorporate elements of scenario analysis and simulations," which "go beyond modeling threats . . . to assess how well private or public organizations would respond to those threats"); Frank Pasquale & Glyn Cashwell, *Four Futures of Legal Automation*, 63 UCLA L. REV. DISCOURSE 26, 28 (2015) (arguing for a scenario analysis in a "time of rapid technological change").

198. Pasquale, *supra* note 26, at 56.

199. *Id.*

rather than "doubling down on the managerialist impulse to quantify costs and benefits."[200]

Orienting AI assessment towards use cases centers the analysis on the distributional and other implications of actual applications on real sets of human beings in concrete contexts. It rejects assessing risk based on what Science and Technology Studies (STS) scholar Donna Haraway calls the "gaze from nowhere."[201] It contests what Professor Sheila Jasanoff argues is risk regulation's "favored . . . type of objectivity"; specifically, it challenges "claims making [that] achieves power by ostensibly detaching knowledge from potentially biased standpoints and from the distortions that any perspective or viewpoint necessarily entails . . . the kind of purification that scientists have historically aimed for in making representations of nature."[202] Instead, a focus on use cases situates the understanding of a system's efficacy, and its risks, in the messy entangled world in which it is located, stabilized, and iteratively repaired, so it can work.

Ideally, then, adopting a use-case orientation moves designers, data scientists, software engineers, and the myriad of others involved in designing, using, justifying, and repairing systems towards Professor Lucy Suchman's concept of "located accountabilities," which replaces "ways of being nowhere while claiming to see comprehensively" with "views from somewhere."[203] These are views built on "partial, locatable, critical knowledges,"[204] and demand that designers take responsibility for what they see and what they "learn how to build."[205]

In sum, the shift in analysis from systems to *use cases* reframes the stakes of AI governance towards a focus on individuals' rights and safety. Stakeholder engagement draws designers' and users' attention to the rights of and impacts on relevant populations. Requiring the involvement of internal subject matter experts responsible for attending to rights within government brings privacy, accessibility, civil rights and security into processes that can shape the design and deployment of systems. Together, these practices center the context in which AI is just one component and directs agencies to question how its introduction will affect agencies' missions and impact the public they serve. By injecting the public's voice and internal rights-oriented expertise into

200. *Id.*

201. Donna Haraway, *Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective*, 14 FEMINIST STUD. 575, 581 (1988).

202. Sheila Jasanoff, *The Practices of Objectivity in Regulatory Science*, *in* SOCIAL KNOWLEDGE IN THE MAKING 307, 309 (Charles Camic, Neil Gross & Michèle Lamont eds., 2011).

203. Lucy Suchman, *Located Accountabilities in Technology Production*, 14 SCANDINAVIAN J. INFO. SYS. 91, 94 (2002) (quoting Donna Haraway, *supra* note 201, at 590).

204. *Id.* at 96 (quoting Donna Haraway, *supra* note 201, at 584).

205. *Id.*

technological design choices the Administration sought to make AI use responsive to public needs and consistent with democratic values.

b) OMB Guidance to Federal Agencies

As described, *supra*, the White House OMB issued guidance to federal agencies on the responsible development and use of AI, and accompanying procurement guidance.[206] The OMB guidance requires agencies to engage affected stakeholders in the design, including risk mitigation choices, of AI systems used by agencies.[207] The federal guidance reflected and built on the Administration's commitments both to public participation and community engagement[208] to developing a more effective set of strategies and tools to support meaningful public participation and community engagement in government policy-making and service design and delivery. At a time of growing distrust in institutions and disaffection with government, the Administration considered these efforts essential to building public trust in government; designing effective, inclusive and accessible policies and services to serve the full public; and aligning government practice with democratic ideals of a government of the people, by the people, and for the people.[209] In addition, it reflected the Biden-Harris Administration's stated commitment to advancing equity across services, policies, and programs, a commitment reflected in the two equity executive orders[210] as well as the effort at modernizing regulatory review to account for the distributional consequences of regulation and to ensure that regulatory initiatives do not unduly burden the disadvantaged.[211]

The guidance established a more demanding set of risk management processes and requirements for government's use of rights-impacting AI, including facial recognition technologies. It centered the engagement of key government experts with responsibility for rights and consultation with and input from affected communities and the public throughout the AI lifecycle.[212]

206. OMB MEMO. M-24-10, *supra* note 101; OMB MEMO. M-24-18, *supra* note 105.

207. OMB MEMO. M-24-10, § 5(v)(B), *supra* note 101, at 22 ("Consult and incorporate feedback from affected communities and the public.").

208. OFF. OF MGMT. & BUDGET, EXEC. OFF. OF THE PRESIDENT, MEMO. NO. M-25-07, BROADENING PUBLIC PARTICIPATION AND COMMUNITY ENGAGEMENT IN THE REGULATORY PROCESS (Jan. 15, 2025), https://bidenwhitehouse.archives.gov/wp-content/uploads/2025/01/M-25-07-Broadening-Participation-and-Engagement.pdf.

209. Methods and Leading Practices for Advancing Public Participation and Community Engagement with the Federal Government, 89 Fed. Reg. 19885, 19886 (Mar. 20, 2024).

210. Exec. Order No. 13,985, 86 Fed. Reg. 7009 (Jan. 20, 2021); Exec. Order No. 14,091, 88 Fed. Reg. 10825 (Feb. 16, 2023).

211. Memorandum on Modernizing Regulatory Review, 86 Fed. Reg. 7223 (Jan. 26, 2021).

212. Deirdre K. Mulligan, Principal Deputy U.S. Chief Tech. Officer, White House Off. of Sci. & Tech. Pol'y, Testimony on the Civil Rights Implications of the Federal Use of

And it further required federal agencies to make the goals, policies, and values embedded in systems and animating their use open and subject to public input.[213]

An interrelated set of requirements build visibility into the AI lifecycle including:

- Data documentation exposing the provenance, the quality and representativeness in relation to purpose, an assessment of its breadth and gaps and how shortcomings of the data have been addressed by the agency or vendor, and if the data is maintained by the Federal Government, whether that it is publicly disclosable as an open government data asset;[214]
- A public use case inventory with accessible documentation in plain language of the system's functionality to serve as public notice of the AI to its users and the general public;[215]
- A requirement for reasonable and timely notice about the use of the AI to those subject to them and a means to directly access any public documentation about it in the use case inventory;[216]
- Notice to individuals when the use of the AI results in an adverse decision or action that specifically concerns them, explanations for such decisions and actions, and if applicable their right to appeal;[217] and,
- A required human fallback and escalation system so impacted individuals can appeal or contest an AI use case's negative impacts[218] and mechanisms for individuals to choose a human alternative where practicable and consistent with law.[219]

A second set of interlocking requirements creates assessments of algorithmic processes that bridge between technical and agency experts and the publics and communities they serve, as well as outside experts:

- Consultation with affected communities, including underserved communities on the design, development, and use of the AI and risk mitigations;[220]

Facial Recognition Technology Before the U.S. Commission on Civil Rights (Mar. 8, 2024), https://www.usccr.gov/files/2024-04/frt-transcript.pdf.

213. OMB MEMO. M-24-10, *supra* note 101.
214. *Id.* §§ 4(d)(ii), 5(c)(iv)(A)(3).
215. *Id.* § 3(a)(iv).
216. *Id.* § 5(c)(iv)(I).
217. *Id.* § 5(c)(v)(D).
218. *Id.* § 5(c)(v)(E).
219. *Id.* § 5(c)(v)(F).
220. *Id.* § 5(c)(v)(B).

- Impact assessments to document the intended purpose for the AI and its expected benefit, potential risks, and quality of the relevant data;[221]
- Testing requirements for performance in real-world contexts;[222]
- Requirements for agencies to identify, assess, and mitigate algorithmic discrimination and harmful bias to ensure that federal government use of AI does not decrease equity or fairness;[223] and,
- Ongoing monitoring and thresholds for periodic human review.[224]

These complement existing requirements such as privacy impact assessments[225] that are intended to bridge the gap between internal experts and the outside world.

Together, these requirements make the values and policy choices built into system designs visible. For those designing and deploying systems, these provide scaffolding for what Phil Agre called "critical technical practice,"[226] in contrast to the formalistic and universalizing tendencies of the field of AI. Agre called for methods and practices of studying, building, and implementing technology that straddled the "craft work of design"[227] and the "reflexive work of critique"[228] to reveal the value choices inherent in technical terminology and design and move towards the useful context specific implementations required for AI to be meaningfully useful. They reflect work in the field of responsible

221. *Id.* § 5(c)(iv)(A).
222. *Id.* § 5(c)(iv)(B).
223. *Id.* § 5(c)(v)(A).
224. *Id.* § 5(c)(iv)(E).
225. DEP'T OF HOMELAND SEC., PRIVACY IMPACT ASSESSMENTS: THE PRIVACY OFFICE OFFICIAL GUIDANCE (June 2010), https://www.dhs.gov/xlibrary/assets/privacy/privacy_pia_guidance_june2010.pdf.
226. Philip Agre, *Toward a Critical Technical Practice: Lessons Learned in Trying to Reform AI*, *in* SOCIAL SCIENCE, TECHNICAL SYSTEMS, AND COOPERATIVE WORK: BEYOND THE GREAT DIVIDE 155 (Geoffrey Bowker, Susan Leigh Star, Les Gasser & William Turner eds., 1997).
227. *Id.*
228. *Id.*

AI that has advanced data and model documentation,[229] auditing,[230] and impact assessments as well as participatory and co-design.[231]

For impacted communities, the system documentation along with requirements for real-world testing and impacted community involvement in risk-mitigation choices, not just design, provide unprecedented opportunities to shape the use of technology. With the use-case orientation, these requirements keep the focus on designing within context and assessing how AI systems affect the outcomes for real people. The notices to negatively impacted individuals and requirements for human fallback and redress create ongoing moments of visibility making space for "technological dramas"[232]—contestation and reexamination of the policies and values baked into the sociotechnical system. These processes acknowledge the inevitability of "algorithmic breakdowns."[233] They create intentional seams that expose the configurability, complexity, and fragility of systems and actively resist the tendency for technological infrastructure to recede into the background.[234]

Finally, Algorithmic Impact Assessments (AIAs) are required to examine the efficacy and risks, including potential bias, in AI systems.[235] They are

229. Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji &Timnit Gebru, *Model Cards for Model Reporting*, FAT* 19: PROCS. OF THE CONF. ON FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY 220 (2019); Ben Hutchinson, Andrew Smart, Alex Hanna, Emily Denton, Christina Greer, Oddur Kjartansson, Parker Barnes & Margaret Mitchell, *Towards Accountability for Machine Learning Datasets: Practices from Software Engineering and Infrastructure*, FACCT '21: PROCS. OF THE 2021 ACM CONF. ON FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY 560, 560–75 (2021); Mark P. Sendak, Michael Gao, Nathan Brajer & Suresh Balu, *Presenting Machine Learning Model Information to Clinical End Users with Model Facts Labels*, 3 NPJ DIGIT. MED. 41 (2020).

230. Inioluwa Deborah Raji, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron & Parker Barnes, *Closing the AI Accountability Gap: Defining an End-To-End Framework for Internal Algorithmic Auditing*, FAT*20: PROCS. OF THE 2020 CONF. ON FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY 33–44 (2020).

231. For a recent overview, see Ned Cooper, Tiffanie Horne, Gillian R. Hayes, Courtney Heldreth, Michal Lahav, Jess Holbrook & Lauren Wilcox, *A Systematic Review and Thematic Analysis of Community-Collaborative Approaches to Computing Research*, CHI '22: PROCS. OF THE 2022 CHI CONF. ON HUMAN FACTORS IN COMPUTING SYS. 1–18 (2022); *see also* SASHA COSTANZA-CHOCK, DESIGN JUSTICE: COMMUNITY-LED PRACTICES TO BUILD THE WORLDS WE NEED (2020).

232. Bryan Pfaffenberger, *Technological Dramas*, 17 SCI., TECH., & HUM. VALUES 282 (1992).

233. Deirdre K. Mulligan & Daniel S. Griffin, *Rescripting Search to Respect the Right to Truth*, 2 GEO. L. TECH. REV. 557, 559 (2018).

234. For a review of the concept of "seamful design," see Sarah Inman & David Ribes, *"Beautiful Seams": Strategic Revelations and Concealments*, CHI '19: PROCS. OF THE 2019 CHI CONF. ON HUMAN FACTORS IN COMPUTING SYS. (2019).

235. OMB MEMO. M-24-10 § 5(c)(iv)(A), (c)(v)(A), *supra* note 101.

intended to ensure that AI systems used by federal agencies do not reproduce existing patterns of discrimination, inhere the prejudice of prior decision makers, or simply reflect the widespread biases that persist in society. Ideally moving away from the litany of existing algorithmic decision-making systems that have been shown to "automate inequality"[236] perpetuating biases of various forms in high-stake consequences including the termination of welfare benefits, granting or denying immigration visas or wrongful imprisonment based on biased AI facial recognition.

Impact assessments, as legal scholar Andrew D. Selbst suggests, encourage those who build systems to think critically about the "potential impacts of a complex project before its implementation, thereby heading off risks before they become too costly to correct."[237] They further create documentation of decisions made during AI systems development to promote accountability for those decisions, as well as to provide useful information for policy interventions to correct for bias down the line.[238] For these reasons, Danielle Keats Citron and Frank Pasquale have advocated for the use of privacy and civil liberties impact assessments when evaluating an AI scoring system's "negative, disparate impact on protected groups, arbitrary results, mischaracterizations, and privacy harms."[239]

At their best, AIAs promote reflexivity, encouraging decisionmakers to document and reflect on their assumptions, their data, and their measurement models at the earliest stages of the development process. Ideally, they act as a *boundary object* offering a shared reference point that enables broader participation in efforts to shape the role of these increasingly consequential technologies in society.

Julie Cohen and Ari Ezra Waldman caution, however, AIAs, as well as other impact or risk assessment frameworks can—like other "regulatory managerial" practices—allow entities to conceal predatory behavior behind a facade of procedural legitimacy.[240] They can also marginalize outside or expert perspectives on the relevant interests at stake (e.g., privacy, racial equity) in favor of "check-box compliance sensibilities."[241] In our own work, therefore, we have foregrounded the potential benefits of using more focused assessment techniques, such as human rights impact assessments (HRIAs), to address

236. *See generally* VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR (2018).

237. Selbst, *supra* note 197, at 122.

238. *Id.* at 118.

239. Danielle Keats Citron & Frank A. Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 26 (2014).

240. Cohen & Waldman, *supra* note 26, at v.

241. *Id.* at vi.

public values more systematically in technical practice.[242] HRIAs focus on impacts on affected communities and center rights.[243] These attributes pull towards a situated consideration of impacts rather than an abstract conversation about system properties. The OMB use-case orientation for AIA similarly foregrounds the institutional and social, along with the technical, and requires agencies to assess how AI systems understood as reconfigurations of work practices or functions[244] will impact the rights and safety of people.

Seen in this light, the AI BoR and the *use-case* framing together reoriented the project and stakes of AI governance and created both a "civic architecture" to support public participation and demanding responsibility and accountability for the impact of technology design and use. Operating within these frames, the OMB requirements provided scaffolding for "critical technical practice"[245] and for ongoing public input and contestation about systems values and impacts throughout the lifecycle.

2. The National Telecommunications and Information Administration (NTIA) and Model Weights

The NTIA public process and report on large AI models with widely available weights—the numerical parameters that comprise an AI model determine its behavior—created a venue to foster the meta-discussion about when and whether it is appropriate to direct aspects of the design of AI models in the service of specific values at all and if so, how to prioritize among values. These are exactly the cross-cutting discussions about whether to engage in a "Governance-by-Design" initiative that we found missing, and without a clear home, in our prior work.[246] The public nature of this process underscored the role government could play in determining whether models with a key property—disclosed weights—were available in the domestic market. The publicness of the process shed light on the possibility that the government would directly shape technology to effect its goal of reducing the risks of AI. The publicness of the process increased the chance that if the

242. Bamberger & Mulligan, *Governance-By-Design*, *supra* note 2, at 764.

243. *See* THE DANISH INST. FOR HUMAN RIGHTS, HUMAN RIGHTS IMPACT ASSESSMENT GUIDANCE AND TOOLBOX 8 (2020), https://www.humanrights.dk/files/media/document/DIHR%20HRIA%20Toolbox_Welcome_and_Introduction_ENG_2020.pdf ("Engagement with rights-holders and other stakeholders is essential . . . . requiring background research and fieldwork, as well as heavily based on the participation of rights-holders other stakeholders").

244. Mulligan & Nissenbaum, *supra* note 150, at 233 (explaining the ways in which reconfiguring how a function is performed can impact rights and values).

245. Kirsten Boehner, Shay David, Joseph 'Jofish' Kaye & Phoebe Sengers, *Critical Technical Practice as a Methodology for Values in Design*, *in* CHI 2005 WORKSHOP: QUALITY, VALUES AND CHOICES, at 1 ("Critical Technical Practice (CTP) is an approach to identifying and altering philosophical assumptions underlying technical practice.").

246. Bamberger & Mulligan, *Governance-By-Design*, *supra* note 2, at 759–70.

government decided to limit the availability of models with widely disclosed weights it would be understood as a modality of regulation rather than a technological inevitability or market decision. It created the conditions for government to be held responsible for the political consequences of shaping technology.[247]

In *Saving Governance-by-Design*, we explained that "[e]xisting governance institutions often lack . . . [the] substantive regulatory capacity—breadth of authority, competence, and vision . . . to support the rational use of technology to govern."[248] We then suggested a number of approaches, including "coordination and input from a range of government actors; and [c]onditioning governance-by-design on multi-stakeholder involvement" that could "broaden the set of values that decision makers must consider, [and] decision makers' capacity to address relevant values."[249] The NTIA process did both.

The President, through the EO 14110, directed NTIA to review the risks and benefits of AI models with widely available weights and develop policy recommendations to maximize those benefits while mitigating the risks.[250] Access to model weights allows fine tuning and the removal of limitations on the model's outputs.

At the President's direction, NTIA sought public input through a Request for Information (RFI) public meetings, and other stakeholder engagement efforts, to ensure that the wide range of public values—from integrity of the scientific process to human rights, to innovation and competition to national security and public safety—were all considered in developing a path forward.[251] This effort is particularly important as government shaping of dual-use technology can have profound implications for the availability and values in applications and services and can hide the government's role in determining the baked-in values and market-availability of technology making it more difficult for the public to practically and legally participate in determining technology's shape and impact.[252]

The NTIA process revealed a wide range of risks and benefits associated with dual-use foundation models with widely available model weights,

247. One of Larry Lessig's key concerns with government shaping of technology is its capacity to obscure government action from the public. LAWRENCE LESSIG, CODE AND OTHER LAWS OF CYBERSPACE 6–8 (1999).

248. Bamberger & Mulligan, *Governance-By-Design*, *supra* note 2, at 759.

249. *Id.* at 760.

250. Exec. Order. No. 14,110, *supra* note 79.

251. NAT'L TELECOMMS. & INFO. ADMIN., DUAL USE FOUNDATION ARTIFICIAL INTELLIGENCE MODELS WITH WIDELY AVAILABLE MODEL WEIGHTS 2 (July 2024), https://www.ntia.gov/sites/default/files/publications/ntia-ai-open-model-report.pdf.

252. Lessig, *supra* note 247 at 6–8.

including public safety, competition, research, among others.[253] It broadened a conversation that had been dominated by concerns that widely available model weights would exacerbate the ability of non-experts to design, synthesize, produce, acquire, or use, chemical, biological, radiological, or nuclear (CBRN) weapons or aid individuals conducting cyberattacks. This process ensured policymakers had a more fulsome picture of the values at stake in prohibiting such models.[254]

The resulting report recommends against restrictions on the wide availability of model weights for dual-use foundation models but recommends that the U.S. government actively collect and evaluate evidence to inform future policy decisions.[255] It evaluates a range of policy approaches, assessing their risks and benefits.[256] And it concludes that, current evidence is not sufficient to definitively determine either that restrictions on such open weight models are warranted, or that restrictions will never be appropriate in the future.[257] The report recommends that the government actively monitor a portfolio of risks that could arise from dual-use foundation models with widely available model weights and take steps to ensure that the government is prepared to act if heightened risks emerge.[258]

The process exemplifies the broad conversation about competing values necessary to wisely enlist design as governance. And precisely because NTIA was directed to explore the implications of technological design choices on a wide range of public rights and values, its recommendations exemplify our overarching design principle of "Design[ing] with Modesty and Restraint to Preserve Flexibility."[259] Because of the complex entangled values at stake, and the limited evidence of specific "marginal risks" from widely disclosed model weights in comparison to withheld model weights, NTIA recommended monitoring of the field rather than directing particular technological deployment choices—namely prohibiting the disclosure of model weights, while recommending the U.S. government undertake activities to collect, evaluate, and if appropriate act on, evidence.[260]

253. NAT'L TELECOMMS. & INFO. ADMIN., *supra* note 251, at 12–34.
254. *Id.*
255. *Id.* at 36.
256. *Id.* at 36–39.
257. *Id.* at 40–47.
258. *Id.*
259. Bamberger & Mulligan, *Governance-By-Design*, *supra* note 2, at 743.
260. NAT'L TELECOMMS. & INFO. ADMIN., *supra* note 251, at 40 ("As of the time of publication of this Report, there is not sufficient evidence on the marginal risks of dual-use foundation models with widely available model weights to conclude that restrictions on model weights are currently appropriate, nor that restrictions will never be appropriate in the future.").

V. CONCLUSION

The Biden-Harris Administration recentered public values in AI governance. It did so through bold public statements that caught the public's imagination, acknowledged the public's experiences, and centered them—their rights and safety, and their communities—in the story of AI's future. It did so through carefully crafted bureaucratic rules that broke with traditional approaches to assessing technology, requiring agencies to assess use cases in all their messy, context-specific complexity. It did so by bringing technologists into government to support the use and governance of AI in collaboration with civil rights, civil liberties, privacy, accessibility, and other subject matter experts. It did so by creating new practices to guide technical design and use that fostered critical reflection and accountability throughout the design and deployment of systems. And it did so by requiring agencies to engage the public in these design and deployment processes and creating visibility throughout them to support public input and accountability. The Administration recognized that how we design, use, and refuse technology is a key way our nation manifests our values. And that getting technology right from the start requires technologists of many sorts to be at the table.

At a moment when rogue technologists are disregarding the rule of law, running roughshod over rights, and weaponizing technology to destroy public institutions, individuals' lives, and democracy, it is foreseeable that some will question the presence of technologists, perhaps suggesting we kick them out of the room. But that would be a mistake for many reasons. Of course, the lawyers currently shredding the government far outnumber the technologists. But more importantly, information and communication technology is a ubiquitous part of government and the institutions it regulates. Delivering robust, rights-respecting services and enforcing the laws that protect the public's rights and safety require technologists to be part of the team.

JUDGEGPT: WHEN PROGRESS MEETS PRECEDENT

Hon. Isabela Ferrari,[†] *Niyati Narang*[††] *& Colleen V. Chien*[†††]

> "I predict that human judges will be around for a while. But with equal confidence I predict that judicial work—particularly at the trial level—will be significantly affected by AI."[1]
>
> –Chief Justice John Roberts

ABSTRACT

This article examines how artificial intelligence technologies are being integrated into the judicial systems of three jurisdictions: Brazil, China, and the United States. Organized around three primary domains of judicial activity—core judicial functions, court management, and interfacing with the public—it provides a comparative perspective on the role of governance, data infrastructure, and local conditions in the realization of AI's promise to improve court efficiency and access.

Responsive to a litigation rate that is among the highest in the world, Brazil has become a leader in judicial AI by leveraging a strong, centralized governance structure and a unified commitment to digital records. Courts are developing proprietary generative AI systems to carry out core functions, utilizing bespoke tools to help with case management, and operating under a sophisticated risk-based governance framework for evaluating use cases. In China, initiatives like the country's "smart court" and Same Type Case Reference system in combination with government aspirations in AI and judicial legitimacy have similarly led to a far-reaching court adoption of automation and AI. This includes national and local initiatives that in certain contexts use intelligent software to find cases, generate draft judgments, and offer streamlined and automated options to the public. In the United States, the judiciary has taken a more cautious and fragmented approach, driven by concerns raised by early experiences with risk assessment tools, an emphasis on due process and judicial autonomy, and a decentralized system of judicial governance. This has so far resulted in a landscape that

---

DOI: https://doi.org/10.15779/Z38086380R

 We thank Yang Ma, Lin Wang, Audrey Im, Alivia Dawson, and the Berkeley Law librarians for excellent research support, and Christopher Hong and the student editors of the Berkeley Technology Law Journal (BTLJ) for their excellent editing and assistance. Prepared for the 2025 BTLJ Spring Symposium, AI Governance at the Crossroads.

† Isabela Ferrari is a Judge in the Federal Court of the Second Region, Brazil, holds a PhD and a Master's degree with honors in Public Law, and was a member of the working group created by the Brazilian National Council of Justice to regulate generative AI in the national judiciary.

†† Niyati Narang is a 3L at Berkeley Law.

††† Colleen V. Chien is a Professor of Law at Berkeley Law and co-director of the Berkeley Center for Law and Technology.

1. C.J. John G. Roberts Jr., 2023 Year-End Report on the Federal Judiciary 6 (Dec. 31, 2023), https://www.supremecourt.gov/publicinfo/year-end/2023year-endreport.pdf.

reflects local innovation, piloting, and experimentation to a greater extent than any top-down mandate. As efforts to centralize and coordinate across the U.S. judiciary take shape, the likelihood of greater technological and procedural legal interoperability—essential for more systematic reform—will also increase.

TABLE OF CONTENTS

I. INTRODUCTION

The rapid evolution of artificial intelligence (AI) presents both opportunities and challenges for the legal system. Much attention has been paid to the uptake of AI by attorneys, and their use of generative AI to perform legal research,[2] draft briefs,[3] and develop evidence at trial.[4] But lawyers are not the only ones turning to AI to navigate caseloads and automate legal tasks. This Article focuses on the ways in which courts and the judiciary are quietly integrating AI within their operations, and in certain contexts and ways, reshaping aspects of the legal system. We focus on three leading jurisdictions—Brazil, China, and the United States—and describe how the courts in each country are approaching the adoption of AI.

While AI can help courts ease backlogs, streamline judicial processes, and enhance consistency in decision-making, it has also raised particular concerns regarding transparency, fairness, and accuracy in addition to the well-rehearsed challenges of government procurement and technical competence. By comparing governance approaches, system-level implementations, and individual use cases, we aim to provide insights into the varying stages of judicial AI adoption and the lessons that emerge from each experience.

---

2. *See The Past, Present, and Future of Legal Research with Generative AI*, THOMSON REUTERS (Feb. 22, 2024), https://legal.thomsonreuters.com/en/insights/white-papers/helping-the-legal-researcher-feel-confident-they-have-done-enough.

3. AI drafts briefs—in many cases—with fake citations. *See* Daniel Wilf-Townsend & Kevin Tobia, *Generative AI and Courts in the United States*, *in* CAMBRIDGE HANDBOOK OF AI AND TECHNOLOGIES IN COURTS (forthcoming 2026) (working paper and abstract), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5243402; Sara Merken, *AI 'Hallucinations' in Court Papers Spell Trouble for Lawyers*, REUTERS (Feb. 18, 2025), https://www.reuters.com/technology/artificial-intelligence/ai-hallucinations-court-papers-spell-trouble-lawyers-2025-02-18/.

4. Natalie Runyon, *Deepfakes on Trial: How Judges Are Navigating AI Evidence Authentication*, THOMSON REUTERS (May 8, 2025), https://www.thomsonreuters.com/en-us/posts/ai-in-courts/deepfakes-evidence-authentication/#:~:text=AI%2Dgenerated%20evidence%20presents%20significant,been%20artificially%20created%20or%20manipulated; *see generally* Abhishek Dalal, Chongyang Gao, Hon. Paul W. Grimm, Maura R. Grossman, Daniel W. Linna Jr., Chiara Pulice, V.S. Subrahmanian & Hon. John Tunheim, *Deepfakes in Court: How Judges Can Proactively Manage Alleged AI-Generated Material in National Security Cases*, 75 U. CHI. LEGAL F. 75 (2024) (offering guidance for judges facing the possibility of AI-generated fake evidence).

For example, in Brazil, a unified commitment to digital records and strong, centralized leadership have provided the foundation for the extensive adoption of AI by the judiciary. Courts are building their own proprietary generative AI systems to meet the demands of a legal system that has the world's highest rate of litigation. A recently-adopted risk-based governance scheme—similar to the E.U. AI Act—prohibits certain uses of AI by the courts (e.g. to predict recidivism based on personality traits) and designates others as high-risk, triggering certain safeguards. In China, by contrast, the country's twin aspirations to both achieve AI excellence if not dominance, and boost the legitimacy and reach of the judiciary, combined with pressures of scale and standardization, have shaped the far-reaching embrace of automation and AI by Chinese courts. National initiatives like the "smart court" movement, in which technology and data play a critical role in service delivery, and the Same Type Case Reference system, which relies on intelligent software systems to find and analyze analogous cases and draft judgments, complement a number of regional efforts that offer "robojudges," self-service mediation options, and predictive bots. A national judicial foundation model promises relief not only to judges, but also legal professionals and members of the public. In contrast to China and Brazil, U.S. courts have taken a more cautious approach to relying on algorithms and AI to produce judicial outputs, particularly in light of concerns raised by early experiences with risk assessment tools and more recent high-profile hallucinations in court filings. The more fragmented and decentralized nature of judicial governance has translated into less uptake of AI systems built specifically for the courts, and more piloting, experimentation, and local innovation across a variety of use cases.

Across all jurisdictions, access to justice remains a challenge. Limited legal literacy, procedural complexity, and geographic disparities have left large swaths of the public without meaningful access to legal help.[5] Against this backdrop, artificial intelligence (AI) and automation have emerged as potential catalysts for changing how courts operate and provide services and assistance to litigants in need. When courts adopt AI—whether to automate routine tasks, assist in core activities, improve the provision of legal aid, or interface with litigants[6]—they not only enhance internal efficiencies but also create new pathways for more inclusive, timely, and transparent access to justice, at least in theory.

Part II provides an overview of how courts operate and includes stylized descriptions of the judicial functions and tasks that have been and are most likely to become AI-augmented. We distinguish between the three major

5. *See infra* Part III.

6. *See* Drew Simshaw, *Interoperable Legal AI for Access to Justice*, 134 YALE L.J. 795, 796 n.5 (2025).

domains of judicial activity: core judicial functions, like legal research and drafting; court management and operations, like case routing; and interacting with the public. Part III provides three jurisdictional vignettes—Brazil, China, and the United States—each organized around the same trio of judicial functions. Part IV highlights cross-cutting insights and themes.

II. THE MULTIFACETED NATURE OF JUDICIAL WORK IN TIMES OF TECHNOLOGICAL TRANSFORMATION

Before we examine how judicial systems around the world are integrating artificial intelligence, we describe the multifaceted nature of judicial work. Behind each visible court opinion or decision lies a complex array of operational and administrative tasks that remain largely invisible to outside observers. One of us has served as a judge for over a decade and can attest that courts are not merely arenas of legal reasoning but dynamic institutions that must be managed, directed, and constantly adapted to shifting caseloads, bureaucratic constraints, and evolving societal needs. Beyond the legal knowledge required to adjudicate cases, judges perform a wide array of managerial and administrative duties that are systematically overlooked in traditional legal education and scholarship yet prove critical to the effective functioning of justice systems.

This "invisible labor" encompasses tasks like managing ever-growing case inventories,[7] organizing and supervising diverse court staff, strategically prioritizing which cases to hear and when, within the margins of procedural discretion, mediating complex expectations among litigants, lawyers, and court personnel, navigating chronically outdated IT infrastructure, and responding to shifting external conditions, including regulatory reforms and rapid technological change.

Much judicial work remains fundamentally manual and "highly repetitive," as judges routinely "spend hours reading long electronic pleading files" that "could be hundreds of pages and usually differ in only a few case-specific features."[8] Similarly, the process of drafting judgments has become "a very laborious and repetitive task for . . . judges, who ha[ve] to collect the relevant

7. A challenge exemplified by India's lower courts, which face a staggering backlog of 40 million cases. *See* Jeremy Barnett, Philip Treleaven, Fredric I. Lederer, Nicolas Vermeys & John Zeleznikow, *JudicialTech Supporting Justice: The Impact of AI and Emerging Technologies on the Judiciary, Courts and Justice* 3 (Univ. of Montr. Fac. of L., Research Paper, 2023), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4597917.

8. *Judicial Systems Are Turning to AI to Help Manage Vast Quantities of Data and Expedite Case Resolution*, IBM (Feb. 4, 2025), https://www.ibm.com/case-studies/blog/judicial-systems-are-turning-to-ai-to-help-manage-its-vast-quantities-of-data-and-expedite-case-resolution.

data and, in the end, repeatedly write almost identical judgments."[9] Judicial officers must therefore function not only as neutral adjudicators applying legal principles, but also as institutional stewards—tasked with safeguarding efficiency, integrity, and meaningful access to justice in increasingly complex and resource-constrained environments.[10]

As Komoda notes, while some judicial functions "are unique and by necessity handled by humans," other more rote responsibilities can consume disproportionate amounts of judicial time and attention.[11] The consequence is clear: "[t]he more time these professionals spend on simple tasks, the less time can be allocated to tasks that require their unique skills and expertise."[12] As the individual productivity of judicial workers wanes, quality and timeliness of legal services across the justice system is systematically undermined.

The emergence of what scholars term "JudicialTech"—defined as "Artificial Intelligence (AI) and emerging technologies' systems for Judges, courts and other forms of dispute resolution"—presents opportunities to address challenges unique to the judiciary.[13] Unlike the broader category of legal technology, JudicialTech specifically focuses on "supporting the Judiciary, enhancing access to Justice, and potentially increasing fairness in the Judicial system."[14]

This technological revolution spans multiple stages of the judicial process: from litigation advice and trial preparation to digital courts and algorithmic decision-support systems.[15] The promise is compelling—artificial intelligence could potentially alleviate administrative burdens, streamline case management, and enhance judicial decision-making through sophisticated data analysis and pattern recognition.

---

9. *Id.*

10. Empirical research reveals that public perceptions of AI integration in judicial systems vary significantly across racial and ethnic groups. Black participants demonstrate notably higher ratings for judicial legitimacy and procedural justice when AI is involved in decision-making compared to White and Hispanic participants, suggesting that historically marginalized communities may view AI as a potential mechanism for reducing judicial bias. This finding challenges assumptions about uniform public resistance to AI in judicial contexts and highlights the importance of considering diverse community perspectives when implementing judicial technologies. *See* Anna Fine, Emily R. Berthelot & Shawn Marsh, *Public Perceptions of Judges' Use of AI Tools in Courtroom Decision-Making: An Examination of Legitimacy, Fairness, Trust, and Procedural Justice*, 15 BEHAV. SCIS. 476, 490 (2025).

11. Jumpei Komoda, *Designing AI for Courts*, 29 RICH. J.L. & TECH. 145, 147 (2023).

12. *Id.* As Komoda further observes, judges and court clerks currently "spend their working hours handling a variety of tasks," where "[s]ome of these are unique and by necessity handled by humans but others are simple, repetitive tasks." This inefficient allocation of human resources "leads to delays and the deterioration of the quality of legal services."

13. Barnett et al., *supra* note 7, at 1.

14. *Id.*

15. *Id.* at 2–3.

Early implementations highlight the potential: courts report that AI assistance allows "judges [to be] relieved of highly repetitive tasks and . . . concentrate on the complex issues," with some systems showing the ability to reduce processing time of cases by over 50%.[16]

However, the integration of AI into judicial systems raises fundamental questions about the nature of judicial authority and the preservation of human judgment in legal proceedings. As courts worldwide experiment with well-established applications like automated document review and online case filing systems as well as more controversial tools such as recidivism prediction algorithms and sentencing support instruments,[17] the judicial community must carefully assess which aspects of their work can be enhanced by technology, and which must be preserved as exclusively human responsibilities.

This technological transformation occurs against the backdrop of broader concerns about algorithmic bias, transparency, and the risk that efficiency gains might come at the cost of procedural fairness or public confidence in judicial institutions. For analytical clarity and considering the diverse fields of AI application in judicial contexts, we organize our examination around three primary domains of judicial tasks where artificial intelligence is making significant inroads:

Core Judicial Functions: This domain encompasses AI tools that directly assist judges in their primary adjudicative responsibilities. These applications include case triage systems that help prioritize urgent matters, predictive jurisprudence analysis that identifies relevant precedents and legal patterns, automated identification of repetitive legal issues that can streamline decision-making processes, and sophisticated document classification systems.[18] Generative AI can also help with a variety of tasks including summarizing and drafting documents, identifying weaknesses in legal arguments, and exploring legal scenarios.

Court Management: The second domain focuses on AI applications that enhance the administrative and managerial aspects of the judicial work that keeps courts functioning effectively. These tools support human resource allocation by analyzing workload patterns and staff performance metrics, strengthen institutional security through advanced monitoring and threat detection systems, improve staff training through personalized learning platforms and competency assessments, and enhance overall administrative efficiency through automation of routine processes and sophisticated data analytics. Courts report that these management-focused AI systems can

16. IBM, *supra* note 8.
17. *See infra* Part III.
18. IBM, *supra* note 8.

significantly reduce administrative burden while improving resource utilization and operational transparency.

Interfacing with the Public: The third domain addresses how AI can bridge the gap between judicial institutions and the citizens they serve. These tools support user services by providing automated guidance and information systems, simplifying communication with litigants through natural language processing and translation services, and facilitating broader access to legal information through intelligent search systems and plain-language explanations of court procedures. Such applications are particularly crucial for addressing access-to-justice concerns, as they can help self-represented litigants navigate complex legal processes and understand their rights and obligations more effectively.

A. UNDERSTANDING ARTIFICIAL INTELLIGENCE IN JUDICIAL CONTEXTS

The term "artificial intelligence" has become increasingly polysemic, encompassing a broad spectrum of technologies, methodologies, and applications that vary significantly in their capabilities, limitations, and appropriate uses within judicial settings. This conceptual ambiguity can lead to both unrealistic expectations and unnecessary skepticism about AI's potential role in supporting judicial work.[19]

For the purposes of judicial analysis, we distinguish between two primary categories of AI technologies, each with distinct characteristics, applications, and implications for judicial practice:

Predictive AI (particularly machine learning systems) represents the more established form of artificial intelligence in judicial contexts. These systems look backward—to historical data—to inform forward-looking decisions. They classify case-related data, identify patterns in judicial reasoning, procedural choices, and case outcomes, and make predictions about future scenarios based on this accumulated information. Predictive AI analyzes such data to discover trends in judicial behavior—such as sentencing patterns, admissibility rulings, or case dispositions—and generate recommendations accordingly.

There is no creative activity in this category of AI; rather, these tools excel at processing vast amounts of information to identify correlations and

19. The Organization for Economic Co-Operation and Development (OECD) defines AI as a "machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments." Komoda, *supra*, note 12, at 148 (citing OECD, *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449 7 (May 21, 2019)). This definition emphasizes AI's capacity to influence decision-making processes while operating within human-defined parameters. *Id.*

statistical relationships that might not be apparent to human reviewers. However, this limitation to historical data analysis for future decision-making creates a significant constraint: predictive AI systems struggle to respond adequately to novel, unprecedented situations that fall outside established patterns.[20]

Generative AI represents a more recent and potentially transformative development in artificial intelligence. While generative systems also analyze historical data, they synthesize this information to create something entirely new—whether text, images, video, or other forms of content.

This creative capacity distinguishes generative AI from its predictive counterpart and opens new possibilities for judicial applications, from drafting documents to providing explanations of legal concepts. However, this same creative activity that enables novel outputs also introduces the risk of "hallucinations"—instances where the AI generates plausible-seeming but factually incorrect or legally unsound content.[21]

B. PREDICTIVE AI IN JUDICIAL PRACTICE

The application of predictive AI in judicial contexts involves analyzing historical patterns of judicial behavior, case outcomes, and procedural decisions to identify trends that can inform future actions. By examining past data to establish patterns of conduct, behavioral tendencies, and decisional frameworks, predictive AI offers numerous potential applications across the judicial spectrum—ranging from low-risk administrative tasks to more complex analytical functions.

The versatility of predictive AI in judicial settings spans activities with varying degrees of risk and complexity. At the lower-risk end of the spectrum, these systems can automate routine procedural tasks and case management functions. More sophisticated applications involve pattern recognition and legal analysis that can support, though not replace, judicial decision-making. Consider several real-world examples of predictive AI implementation:[22]

- **Low-Risk Applications:** Identification of *lis pendens*, issue preclusion (collateral estoppel), or time-barred lawsuits represents perhaps the most straightforward application of predictive AI in judicial contexts. These systems can automatically flag potential procedural problems by

20. On the operation of predictive AI, see generally Jürgen Schmidhuber, *Deep Learning in Neural Networks: An Overview*, 61 NEURAL NETWORKS 85–117 (2015).

21. For an accessible technical explanation of how generative artificial intelligence operates, see ANDREJ KARPATHY, *Deep Dive Into LLMs Like ChatGPT*, at 1:03:45 (YouTube, Feb. 5, 2025), https://youtu.be/7xTGNNLPyMI?si=7ODBzt1qHF5zDpoj.

22. For the application of artificial intelligence in the judiciary, with real cases and examples, see ISABELA FERRARI, DISCRIMINAÇÃO ALGORÍTMICA E PODER JUDICIÁRIO [ALGORITHMIC DISCRIMINATION AND THE JUDICIARY] 75–95 (Emais ed., 2023).

comparing new filings against existing case databases. Similarly, electronic judicial attachments and asset research can be significantly enhanced through AI systems that search and cross-reference financial databases more efficiently than manual processes.

- **Moderate-Risk Applications:** Classification of lawsuits—defining appropriate procedural processes and grouping similar cases—requires more sophisticated analysis but offers substantial efficiency gains. AI systems can analyze case characteristics, legal claims, and procedural requirements to recommend appropriate case tracks and identify similar matters for consolidated handling. Additionally, recommendation of applicable laws and binding precedents can assist judicial officers by identifying relevant legal authorities, though the final determination of legal applicability remains a judicial function.[23]
- **High-Risk Applications:** Identification of similar cases to use their rulings as models represents a more complex application that begins to approach core judicial functions. These systems can analyze legal issues, factual patterns, and case outcomes to suggest relevant precedents and analogical reasoning frameworks.[24]

C. GENERATIVE ARTIFICIAL INTELLIGENCE IN JUDICIAL CONTEXTS

Unlike predictive AI systems, which analyze historical data to identify behavioral patterns and make recommendations based on established precedents, generative artificial intelligence represents a fundamentally different technological paradigm. Generative AI systems autonomously produce novel content—including texts, summaries, transcriptions, legal analyses, and even codes—based on large-scale language models trained on vast corpora of existing material. These systems simulate a form of machine

23. While this human-centric approach may seem intuitive to legal professionals, recent empirical research reveals that it reflects genuine public sentiment rather than mere regulatory compliance. In a comprehensive study examining public perceptions across racial and ethnic groups, researchers found consistent emphasis on preserving human agency in judicial decision-making, with participants explicitly articulating that "[j]udges should use artificial intelligence to help guide their decisions but they shouldn't base their final decision on the AI's suggested action." Fine et al., *supra* note 10, at 489. This finding suggests that successful AI implementation in judicial contexts must navigate not only technical and legal constraints, but also deeply held public expectations about the irreplaceable role of human judgment in legal proceedings. *Id.*

24. At the highest end of complexity, some jurisdictions have experimented with drafting decisions based on previous rulings—essentially using predictive AI to generate judicial opinions by synthesizing patterns from historical decisions; however, such tools have increasingly been abandoned as generative artificial intelligence has gained prominence, offering more sophisticated approaches to judicial writing assistance.

creativity that extends beyond pattern recognition to genuine content creation, finding increasing application within the legal sector and beyond.

The creative capacity of generative AI emerges from its ability to synthesize information from multiple sources and produce entirely new outputs rather than simply retrieving existing materials. When prompted to generate content, these systems do not merely locate and return the most relevant existing document, as traditional search engines might. Instead, they analyze patterns across their training data and create novel combinations that respond to specific prompts while maintaining coherence and relevance. This synthetic process produces outputs that are inherently stochastic—probabilistic in nature—meaning that identical prompts may yield different results across multiple interactions, as the system continuously recombines learned patterns in novel ways.

The stochastic nature of generative AI reflects the mathematics underlying these systems. Rather than following deterministic rules that produce identical outputs for identical inputs, generative models sample from probability distributions learned during training, introducing controlled randomness into the generation process. This probabilistic approach enables creativity and prevents mechanical repetition but simultaneously introduces unpredictability that poses particular challenges for judicial applications where consistency and reliability prove paramount.

Current applications of generative AI within judicial contexts span both general-purpose tools and specialized legal technologies. General platforms such as ChatGPT, Claude, Gemini, and similar systems offer broad capabilities applicable to diverse legal tasks, while more specialized tools focus on particular judicial functions such as creating legal knowledge repositories, generating procedural diagrams, and summarizing judicial opinions. These specialized applications often incorporate domain-specific training data and fine-tuning to enhance their relevance and accuracy within legal contexts.[25]

The deployment of generative AI within judicial systems occurs across two primary domains, each presenting distinct opportunities and challenges. Private use by individual judges encompasses applications such as drafting assistance for opinions and judicial decisions, analysis of case details and legal arguments, exploration of alternative legal scenarios and precedential frameworks, summarization of voluminous legal documents, identification of inconsistencies within witness testimony, and assistance in formulating questions for expert witnesses. These individual applications offer immediate

25. In the Brazilian judiciary, various training programs have been offered to judges on the use of generative AI in legal writing. Among these, courses taught by Hon. Judge George Marmelstein stand out for their focus on practical applications.

productivity benefits but raise concerns about consistency across judicial officers and adequate oversight of AI-generated content—issues further explored in later sections addressing hallucinations, regulatory safeguards, and the need for human review (see *infra* Section II.C and Part IV).

Institutional deployment of generative AI requires more systematic approaches to implementation, training, and governance. General institutional adoption demands comprehensive education programs to ensure judicial officers understand both capabilities and limitations of these technologies. Specialized systems designed for particular legal tasks represent emerging initiatives, exemplified by projects such as Assis,[26] developed by the Rio de Janeiro State Court, which provides tailored assistance for specific judicial functions. Such institutional implementations require careful attention to cost considerations, data governance agreements protecting sensitive information, and systematic quality control mechanisms.

Generative AI demonstrates particular strengths in several areas relevant to judicial work. These systems excel at improving existing texts through correction, summarization, and stylistic adaptation, enabling judges to refine drafts and enhance clarity of judicial opinions. Their capacity for comparing and contrasting different legal concepts facilitates analysis of competing arguments and identification of relevant distinctions. The ability to suggest alternative approaches and creative solutions can assist judicial officers in exploring novel legal theories or considering unconventional remedies. Additionally, these systems can construct coherent narratives and timelines from complex factual scenarios, potentially assisting in case organization and presentation.

However, generative AI introduces unique challenges that compound the difficulties already identified with predictive systems. The creative capacity that enables novel content generation simultaneously creates the risk of hallucinations—the production of plausible-seeming but factually incorrect or legally unsound content.[27] These hallucinations emerge from various sources, including insufficient or biased training data, incorrect assumptions embedded within model architecture, design priorities that emphasize pattern-based content generation over factual accuracy, adversarial manipulation by bad

26. *ASSIS-Assistente de Inteligência Artificial Generativa* [*ASSIS-Generative Artificial Intelligence Assistant*], PODER JUDICIÁRIO: ESTADO DO RIO DE JANEIRO [JUDICIAL BRANCH: STATE OF RIO DE JANEIRO], https://www.tjrj.jus.br/magistrado/servicos/assis/o-projeto.

27. This risk has already led to real-world concerns in judicial systems. *See, e.g.*, Robert Booth, *High Court Tells UK Lawyers to Stop Misuse of AI After Fake Case-Law Citations*, GUARDIAN (June 6, 2025), https://www.theguardian.com/technology/2025/jun/06/high-court-tells-uk-lawyers-to-urgently-stop-misuse-of-ai-in-legal-work?CMP=share_btn_url.

actors seeking to corrupt outputs, and fundamental limitations of current AI technology.

The interactive nature of generative AI systems creates additional pathways for problematic outcomes through user misuse. Discrimination may arise not merely from training data or algorithmic design, but from patterns of user interaction,[28] including biased prompting strategies, selective information-seeking behavior, or unconscious reinforcement of existing prejudices through iterative questioning. The flexibility of these systems makes them susceptible to jailbreaking attempts wherein users circumvent ethical restrictions and safety measures to elicit inappropriate or harmful content.

Mitigating these risks within judicial contexts[29] requires multifaceted approaches that combine technical solutions with procedural safeguards. Specific prompting strategies can enhance accuracy and reduce hallucination risks by providing clear context, explicit constraints, and verification requirements. Ethical alignment mechanisms work to prevent jailbreaking attempts and ensure outputs conform to professional and legal standards. Grounding techniques connect AI outputs to verifiable sources of information, enabling validation and reducing reliance on potentially fabricated content. Systematic review processes ensure human oversight of AI-generated content before incorporation into official judicial work.[30] Finally, emerging regulatory frameworks seek to establish appropriate boundaries and accountability mechanisms for AI use within judicial systems.

The contemporary judicial landscape presents fertile ground for both predictive and generative AI applications, with courts worldwide experimenting with diverse implementations across the spectrum of judicial functions. The experiences of leading jurisdictions provide valuable insights

28. Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 674, 715–23 (2016).

29. For a critical examination of the legal and ethical implications that arise when GenAI is used in judicial rulings and their underlying rationale, see generally David Uriel Socol de la Osa & Nydia Remolina, *Artificial Intelligence at the Bench: Legal and Ethical Challenges of Informing—Or Misinforming—Judicial Decision-Making Through Generative AI*, 6 DATA & POL'Y e59 (2024).

30. The stakes of inadequate oversight extend far beyond technical errors—they touch the very core of judicial legitimacy and public trust. In a recent multi-ethnic study examining public perceptions of AI in judicial decision-making, researchers found that public anxiety about AI in judicial contexts centers not merely on accuracy concerns, but on the irreversible consequences of algorithmic mistakes. Fine et al., *supra* note 10, at 489. One participant's warning captures this visceral concern: "Judges using AI in decision-making need to be extremely careful that they don't make a mistake and ruin someone's life by misusing AI." *Id.* This sentiment reflects a deeper understanding that judicial errors involving AI carry unique reputational and systemic risks that could undermine confidence in the entire justice system, making robust human oversight not just a technical necessity but a prerequisite for maintaining judicial legitimacy. *See id.*

into both the transformative potential and persistent challenges associated with AI integration within judicial systems, offering lessons that can inform responsible development of these technologies across diverse legal contexts.

III. FROM CODE TO COURTROOM: NATIONAL APPROACHES TO AI IN THE JUDICIARY

This Part surveys the experiences of the Brazilian, Chinese, and American judiciaries with AI. Of note, we dedicate more space to the Brazilian case by drawing upon the personal experiences of one of us as a sitting federal judge who has witnessed firsthand how artificial intelligence has been considered, negotiated, and deployed in that jurisdiction. In contrast, our profiles of China and the United States draw from publicly available documents and secondary literature. While each country's profile is organized around the same core aspects of the judicial experience with AI, readers should bear in mind the asymmetry of underlying sources and evidence we draw upon.

A. BRAZIL: MODEL IN THE MAKING

One of the highest litigation rates in the world, an early and unified commitment to digital records, and strong, centralized governance of AI are some of the factors that have led Brazil to be recognized as a leader in the use of AI in the judiciary.

1. A Landscape of High Litigation: Structural Pressures on the Judiciary

Brazil's judiciary is a unified system with five branches—Federal, State, Labor, Electoral, and Military—each operating independently under the 1988 Constitution. At the top of the judicial hierarchy, the Supreme Court (STF) functions as the constitutional court responsible for safeguarding fundamental rights, followed by the Superior Court of Justice (STJ), which ensures uniform interpretation of federal law across jurisdictions. Administrative oversight is exercised by the National Council of Justice (CNJ), a non-adjudicative constitutional body tasked with coordinating national judicial policy, monitoring court performance, and regulating the use of technology. While the CNJ supervises all courts and judges, it does not oversee the STF itself, which remains institutionally independent.

Despite centralized governance by the CNJ and superior courts, first-instance judges in Brazil exercise significant administrative autonomy. Each presides over a *vara*—a unit that functions as both courtroom and administrative office—managing staff, dockets, hearings, and procedural routines. This discretion has enabled judges to pilot digital tools, streamline workflows, and develop local initiatives to improve access to justice. The result

is a hybrid model: nationally coordinated but operationally decentralized, fostering responsiveness and innovation while also posing coordination challenges.

These dynamics are compounded by structural pressures. Broad constitutional guarantees of access to justice,[31] including state-funded legal aid, have led to extremely high litigation rates. As of 2022, fewer than 7,000 public defenders served a population of over 210 million[32]—forcing courts to rely on private attorneys (*advogados dativos*).[33] In 2021, nearly 30% of all cases benefited from state-funded legal assistance, underscoring the system's scale and social role.[34] Moreover, Brazil's legal profession is vast: 1.4 million licensed lawyers as of 2024, or one for every 152 people—the highest per capita ratio in the world.[35] This density increases demand for judicial services and contributes to procedural complexity, reinforcing the need for scalable, technology-driven solutions. Together, these features—judicial autonomy, universal access guarantees, litigation overload, and professional density—create an environment where AI is not only desirable but increasingly essential to ensure institutional sustainability and procedural fairness.

While Brazil's constitutional guarantees and institutional mechanisms have laid a strong foundation for access to justice, the system must operate under exceptionally demanding conditions. Brazil is widely recognized for having one of the highest litigation rates in the world. As of 2023, there were approximately 83.8 million active legal proceedings across 91 courts

31. CONSTITUIÇÃO DA REPÚBLICA FEDERATIVA DO BRASIL [CONSTITUTION OF THE FEDERATIVE REPUBLIC OF BRAZIL] art. 5, cl. XXXV (1988) ("The law shall not exclude from the assessment of the Judiciary any injury or threat to a right."); *id.* art. 5, cl. LXXIV ("The State will provide full and free legal assistance to those who prove insufficient resources.").

32. DIOGO ESTEVES, CLEBER ALVES & ANDRÉ CASTRO, *National Report: Brazil*, 7 (International Legal Aid Group Conference, 2017), https://clp.law.harvard.edu/wp-content/uploads/2023/06/Brazil-National-Report-ILAG-Conference-2023.pdf.

33. *Pro Bono Practices and Opportunities in Brazil*, LATHAM & WATKINS LLP 91, 93–94 (Feb. 21, 2016), https://www.lw.com/admin/Upload/Documents/Global%20Pro%20Bono%20Survey/pro-bono-in-brazil.pdf.

34. Skye.Tan.22, *Brazilian Legal Clinics Work to Promote Effective Access to Justice*, OFF. BLOG UCL STUDENT PRO BONO COMM. (Dec. 8, 2023), https://reflect.ucl.ac.uk/access-to-justice/2023/12/08/brazilian-law-clinics-work-to-promote-effective-access-to-justice/; *Justiça em Números 2022* [*Justice in Numbers 2022*], CONSELHO NACIONAL DE JUSTIÇA [NATIONAL COUNCIL OF JUSTICE] 1, 115 (2022), https://www.cnj.jus.br/wp-content/uploads/2022/09/justica-em-numeros-2022-1.pdf.

35. Renato Souza, *O País Com Mais Advogados* [*The Country with the Most Lawyers*], CORREIO BRAZILIENSE [BRAZILIAN MAIL] (Aug. 11, 2024), https://www.correiobraziliense.com.br/politica/2024/08/6917908-o-pais-com-mais-advogados.html. The daily updated number of lawyers in Brazil can be found at *Quadro da Advocacia*, ORDEM DOS ADVOGADOS DO BRASIL NACIONAL [NATIONAL ORDER OF ATTORNEYS OF BRAZIL], https://www.oab.org.br/institucionalconselhofederal/quadroadvogados.

nationwide[36]—a volume that places extraordinary pressure on judicial personnel and infrastructure. Settlement rates remain low, at around 12%,[37] meaning that most disputes must be resolved through formal judicial decisions. On average, each judge is responsible for closing over 2,000 cases per year, which translates to approximately 8.6 final rulings (*sentenças*) per working day.[38] In addition to these final decisions, judges must also issue a high number of interlocutory rulings throughout the life of each proceeding, addressing procedural motions, evidentiary requests, and other case-management tasks. This cumulative workload has intensified the demand for technological solutions that can support caseflow management, reduce administrative burdens, and enhance the efficiency of judicial decision-making.

2. Building Digital Justice: Data Consolidation and Policy Coordination

Brazil's digital transformation began in 2006 with the adoption of electronic case processing across its judiciary.[39] While implementation initially varied across branches and jurisdictions—resulting in a fragmented technological landscape—by 2021, 97.2% of all new cases were filed electronically, reflecting a rapid and near-universal shift to digital justice.[40] This process generated not only operational efficiency but also an unprecedented volume of structured procedural data, laying the foundation for institutional AI deployment.

To harness this data potential, the National Council of Justice (CNJ) developed CODEX, a centralized data lake that integrates structured and unstructured data from multiple case management systems into a unified national repository.[41] Created in partnership with the Court of Justice of Rondônia, CODEX consolidates over 337 million legal proceedings, including full-text documents, metadata, and procedural information.[42] It supports advanced analytics, regulatory design, and model training, and functions as the

36. *Justiça em Números 2022* [*Justice in Numbers 2022*], *supra* note 34, at 15.

37. *Justiça em Números 2024* [*Justice in Numbers 2024*], at 252.

38. *Id.* at 20.

39. Law No. 11.419/2006 authorized the use of electronic records in judicial proceedings and established the legal foundation for nationwide digitalization.

40. Luciana Otoni, *Justiça em Números 2022: Processos Eletrônicos Alcançam 97,2% das Novas Ações* [*Justice in Numbers 2022: Electronic Processes Reach 97.2% of New Cases*], CONSELHO NACIONAL DE JUSTIÇA [NATIONAL COUNCIL OF JUSTICE] (Sep. 16, 2022), https://www.cnj.jus.br/justica-em-numeros-2022-processos-eletronicos-alcancam-972-das-novas-acoes.

41. *Platforma Codex* [*Codex Platform*], CONSELHO NACIONAL DE JUSTIÇA [NATIONAL COUNCIL OF JUSTICE], https://www.cnj.jus.br/sistemas/plataforma-codex/.

42. *Codex-Público* [*Public Codex*], METABASE, https://metabase.ia.pje.jus.br/public/dashboard/d4c8362c-4150-4359-96c9-b5cbf1f64f15.

technical backbone of Brazil's AI ecosystem, enabling scalable and responsible AI integration across courts.

CODEX represents a rare instance of judicial data infrastructure at national scale—positioning Brazil among the few jurisdictions in the world with the institutional capacity to develop AI solutions grounded in real case data and legal reasoning patterns. One of its major achievements lies in addressing Brazil's historically fragmented digital justice environment, where numerous electronic case management systems were developed in a decentralized, often bottom-up fashion.

This fragmentation posed significant interoperability challenges—both within the same court system (e.g., between first and second instances) and across jurisdictions, including communication with superior courts such as the Superior Court of Justice (STJ) and the Supreme Court (STF).[43] CODEX's centralized architecture enables seamless data integration across these disparate systems, facilitating jurisdictionally consistent AI governance and nationally coordinated innovation.

This transition from digital case management to strategic data governance was supported by strong institutional leadership. During his tenure as President of the Supreme Court (STF) and the CNJ (2018–2020), Justice José Antonio Dias Toffoli convened judges and technical experts to develop a national vision for judicial AI. The result was the 2019 Handbook on Artificial Intelligence in the Brazilian Judiciary,[44] which outlined key principles—transparency, human oversight, fairness—and documented 14 early-stage AI initiatives.

Building on this foundation, the CNJ issued Resolution No. 332/2020,[45] one of the world's first regulatory frameworks specifically designed for judicial

43. Interoperability in the context of Brazil's electronic judicial process refers not merely to the ability of systems to exchange data, but to a broader set of technical, legal, organizational, and semantic conditions that allow integrated and functional communication between different platforms. For an in-depth empirical study on the interoperability challenges faced by Brazilian courts—including intra–institutional gaps and user–centered functionality—see Carlos Renato Cunha, Cesar Antonio Serbena, Cesar Felipe Bolzani, Edna Torres Câmara, Gustavo Vieira Vilar Garcia, Nayara de Camargo Pinto, Priscila da Silva Barbosa & Renê Chiquetti Rodrigues, *Pesquisa Nacional: Interoperabilidade dos Sistemas de Processo Eletrônico no Brasil* [*National Survey: Interoperability of Electronic Process Systems in Brazil*], CNPQ (2018), https://www.cnj.jus.br/wp-content/uploads/2018/09/d22fe00c12cda4219b4876efb44bfc42.pdf.

44. *Inteligência Artificial: No Poder Judiciário Brasileiro* [*Artificial Intelligance: In the Brazilian Judiciary*], CONSELHO NACIONAL DE JUSTIÇA [NATIONAL COUNCIL OF JUSTICE] (2019), https://bibliotecadigital.cnj.jus.br/jspui/bitstream/123456789/98/1/Intelig%c3%aancia%20Artificial%20no%20Poder%20Judiciario%20Brasileiro.pdf.

45. CONSELHO NACIONAL DE JUSTIÇA RESOLUÇÃO [NATIONAL COUNCIL OF JUSTICE RESOLUTION] NO. 332 (2020), https://atos.cnj.jus.br/atos/detalhar/3429.

AI. Inspired in part by international debates such as *State v. Loomis* in the United States, the resolution addresses core risks of opacity, bias, and lack of explainability, and sets forth safeguards to align AI systems with constitutional guarantees—particularly human dignity, liberty, due process, and equality.[46] It prohibits the use of AI in criminal sentencing and risk assessments, strongly reinforces the principle of human oversight, and mandates that technology serve strictly as a non-decisional auxiliary tool.[47]

The resolution also emphasizes cybersecurity and data protection, recognizing the sensitivity of judicial records.[48] In doing so, it positions Brazil as a pioneer in rights-based AI governance, embedding normative caution and institutional accountability at the center of its innovation strategy.[49]

3. A New Regulatory Framework: From CNJ Resolution No. 332/2020 to CNJ Resolution No. 615/2025

As generative AI tools began to permeate judicial practice, the need for an updated regulatory framework became evident. In response, the current President of the Brazilian Supreme Court, Chief Justice Luís Roberto Barroso, convened a multidisciplinary working group under the auspices of the National Council of Justice (CNJ). The group was tasked with updating the regulatory parameters first established by Resolution No. 332/2020, in light of the rapid and diverse forms of generative artificial intelligence ("GenAI") adoption already taking place across the judiciary.[50]

This widespread adoption unfolded along two parallel paths. First, as described below, judges began experimenting individually with general-purpose tools such as ChatGPT and Claude, applying them to auxiliary tasks like summarizing documents, transcribing hearings, and drafting preliminary analyses. As interest grew—fueled by increasing workloads and oversubscribed judicial training sessions—usage expanded informally across courts. Simultaneously, even before new regulation was in place, some courts initiated the development of institutional GenAI systems, customized to specific judicial workflows. The coexistence of unregulated individual use and early institutional experimentation underscored the urgency of a revised framework and provided critical input to the working group's deliberations.

Active throughout 2024 and the beginning of 2025, the working group brought together representatives from the judiciary, public defenders' offices,

46. *Id.* arts. 2, 7.
47. *Id.* arts. 17(I)–(II), 23 §§ 1–2.
48. *Id.* arts. 13–16.
49. *Id.* arts. 9, 10, 25–27.
50. CONSELHO NACIONAL DE JUSTIÇA PORTARIA [NATIONAL COUNCIL OF JUSTICE ORDINANCE] NO. 338 (2023), https://atos.cnj.jus.br/atos/detalhar/5368.

prosecutors' offices, the legal profession, academia, technical experts, and leading specialists in data protection.[51] Its mission was to align judicial AI governance with emerging technological realities—especially those introduced by generative AI—and to ensure that innovation proceeds without compromising constitutional guarantees.

The result of this effort was CNJ Resolution No. 615/2025,[52] issued on March 11, 2025, which introduces a comprehensive and binding framework for the development, deployment, and oversight of AI systems within the Brazilian judiciary. While CNJ Resolution No. 332/2020 remained in force during a 120-day transition period,[53] the new resolution significantly broadens the regulatory scope in both substance and structure.

At its core, the resolution aims to foster responsible technological innovation in judicial services without compromising constitutional guarantees. It affirms that all AI systems must comply with fundamental principles enshrined in the Brazilian Federal Constitution—particularly the protection of human dignity, the prohibition of discrimination, and the right to due process.[54] Importantly, it mandates that all AI-assisted activities remain subject to human oversight[55] and that AI outputs must be explainable, verifiable, and open to challenge.[56]

A key innovation lies in the introduction of a tiered, risk-based classification system. AI tools are categorized as low, medium, or high risk depending on their potential to interfere with individual rights or procedural guarantees. This classification determines the level of transparency, supervision, and auditability required.[57] High-risk systems, for instance, are subject to stricter controls and must undergo regular review and lifecycle documentation.[58]

Recognizing the distinctive challenges posed by generative AI, the resolution sets forth specific rules governing its use. These include requirements for output transparency, mandatory human review prior to the integration of AI-generated content into judicial decisions or draft opinions,

51. One of the authors, Federal Judge Isabela Ferrari, served as a member of the multidisciplinary working group convened by the National Council of Justice (CNJ) to draft Resolution No. 615/2025.

52. CONSELHO NACIONAL DE JUSTIÇA RESOLUÇÃO [NATIONAL COUNCIL OF JUSTICE RESOLUTION] NO. 615 (2025) [hereinafter CNJ Res. No. 615], https://atos.cnj.jus.br/atos/detalhar/6001.

53. *Id.* art. 47.

54. CNJ Res. No. 615, *supra* note 52.

55. *Id.* arts. 2(V), 3(VII), 15–18.

56. *Id.* arts. 3(II), 13(VII), 22 § 3.

57. *Id.* arts. 9, 11, Anexo [Appendix].

58. *Id.* arts. 11 § 1–2, 13(IV), 14 § 1, 17 § 1.

and an explicit prohibition against relying solely on AI outputs as the basis for legal rulings. Importantly, the resolution does not prohibit the automation of supporting functions—such as document summarization, transcription, or drafting assistance—but draws a firm line against delegating the core judicial function: the authoritative act of legal interpretation and decision-making. Judges may not use generative AI to determine legal outcomes or to ask what should be decided; the final reasoning and judgment must remain both formally and substantively human.[59] These measures are designed to preserve judicial reasoning not merely as a procedural formality, but as an intrinsically human exercise grounded in deliberation, responsibility, and legal authority.

In addition to these safeguards, the resolution outlines a list of expressly prohibited uses.[60] AI systems may not be used to predict criminal behavior or the likelihood of recidivism based on personality traits or behavioral profiling. Nor may they classify or rank individuals in ways that affect access to legal rights or operate without meaningful human intervention. These prohibitions reflect a clear stance against algorithmic determinism in contexts that demand human judgment and legal nuance.

To support institutional accountability, the resolution requires courts to establish internal governance mechanisms for AI oversight.[61] These include periodic audits of high–risk systems and detailed documentation of each system's lifecycle—from initial development and data training to deployment, monitoring, and eventual deactivation. Such requirements aim to ensure traceability, enable institutional learning, and mitigate the risks associated with opaque or poorly understood systems.

Transparency is further enhanced through the mandatory registration of all judicial AI tools on Sinapses,[62] a centralized national platform maintained by the National Council of Justice. This registry functions as both a compliance tool and a strategic database, helping to avoid technological fragmentation, promote interoperability, and enable coordinated supervision across the judiciary.[63]

Finally, the resolution addresses the cultural and educational conditions necessary for safe and effective AI adoption. It mandates continuous training programs for judges and court personnel, with particular emphasis on algorithmic bias, data protection, and the ethical limits of automation.[64] By

59. *Id.* arts. 19 § 3(II), 6, 20(IV).

60. *Id.* art. 10.

61. *Id.* art. 12(III).

62. *Id.* art. 24.

63. *Sinapses Platform*, CONSELHO NACIONAL DE JUSTIÇA [NATIONAL COUNCIL OF JUSTICE], https://www.cnj.jus.br/sistemas/plataforma-sinapses/.

64. CNJ Res. No. 615, *supra* note 52, arts. 3(VIII), 16(VII), 20(III), 19 § 3(I), 5.

incorporating education into the regulatory framework, the resolution seeks not only to build institutional capacity, but also to promote critical awareness of the roles and limitations of AI in legal decision-making.

The following sections present examples of how different forms of AI—both predictive and generative—have been integrated into the Brazilian judiciary, spanning court management, citizen-facing services, and core judicial functions.

a) Integration of Generative AI into Core Judicial Functions

As described earlier, the use of generative AI by the Brazilian judiciary has taken two distinct paths. First, judges have been experimenting individually with general-purpose tools such as ChatGPT, Claude, and Gemini, often through personal subscriptions. These individual adoptions have been informal and exploratory, with judges using these tools for a variety of auxiliary functions.[65] This initial wave of adoption was driven both by growing interest among judges and by official training programs organized by judicial schools.[66] These courses—quickly oversubscribed—reflect a strong demand for practical, constitutionally grounded guidance on the opportunities and risks posed by generative AI.

65. These include summarizing documents and accelerating the review of lengthy case files, converting oral testimonies and hearings into text, comparing witness statements and assisting in the detection of inconsistencies, drafting questions for witness examination and supporting the formulation of strategic inquiries, conducting legal research and retrieving applicable case law or doctrinal references. While hallucinations and factual inaccuracies remain a concern, ongoing improvements aim to reduce these limitations.

66. For examples of institutional training, see the 2025 course "Inteligência Artificial Generativa Aplicada ao Judiciário" (Generative Artificial Intelligence Applied to the Judiciary), organized by the Center for Judicial Studies (Centro de Estudos Judiciários–CEJ) of the Federal Justice Council (Conselho da Justiça Federal–CJF). The program included both basic and advanced modules on prompt engineering, data protection, and the ethical use of generative AI in judicial practice. *See Inteligência Artificial Generativa Aplicada ao Judiciário* [*Generative Artificial Intelligence Applied to the Judiciary*], CENTRO DE ESTUDOS JUDICIÁRIOS: CONSELHO DA JUSTIÇA FEDERAL [CENTER FOR JUDICIAL STUDIES: FEDERAL JUSTICE COUNCIL] (2025), https://www.cjf.jus.br/cjf/corregedoria-da-justica-federal/centro-de-estudos-judiciarios-1/eventos/eventos-cej/2025/IAgaJud-2025-pres.
Another relevant example is the course "Escrita Jurídica com o ChatGPT: Teoria e Prática" (Legal Writing with ChatGPT: Theory and Practice), held by the School for the Judiciary (Escola Superior da Magistratura–ESMA). Taught by federal judge George Marmelstein, the course covered legal writing with AI support, prompt safety, and persuasive writing strategies using ChatGPT, emphasizing cautious and informed adoption of generative AI in judicial routines. *See* Marcus Vinícius, *O Uso de IA Nas Atividades dos Magistrados é Tratado No Curso 'Escrita Jurídica Com o ChatGPT'* [*The Use of AI in Magistrates' Activities is Addressed in the Course 'Legal Writing with ChatGPT'*], PODER JUDICIÁRIO: TRIBUNAL DE JUSTIÇA DA PARAÍBA [JUDICIAL BRANCH: PARAÍBA COURT OF JUSTICE] (Apr. 8, 2024), https://www.tjpb.jus.br/noticia/o-uso-da-ia-nas-atividades-dos-magistrados-e-tratado-no-curso-escrita-juridica-com-o-chatgpt.

Second, courts have started to institutionalize the use of these technologies by developing and implementing customized systems designed to support judicial workflows. These systems include court-sanctioned tools—whether developed internally or adopted through official channels—that are integrated into judicial workflows and governed by constitutional and procedural constraints.

Among the most significant developments are proprietary generative AI systems built by the courts themselves. These tools reflect a transition from experimentation to structured, rule-bound deployment, with the goal of improving procedural efficiency while ensuring transparency and preserving judicial authority, and include:

- ASSIS (TJRJ):[67] Developed by the Rio de Janeiro State Court (TJRJ), ASSIS—short for Assistente de Inteligência Artificial com Soluções de Sentença—is a generative AI assistant designed to support judges in drafting decisions and opinions. It integrates with the court's case management system, allowing judges to request AI-generated drafts and legal summaries based on case files and existing jurisprudence.
- ChatJT (Labor Courts):[68] The Labor Justice system has launched ChatJT, a generative AI system trained on case law from the first instance through the Superior Labor Court (TST). It is designed to promote coherence in legal reasoning and to assist judges in producing rulings aligned with prevailing precedents.
- Logos (STJ):[69] The Superior Court of Justice (STJ) introduced Logos, a generative engine that aids in the preparation of draft rulings and case assessments. Its objective is to increase both the speed and uniformity of judicial decisions, especially in repetitive or high-volume matters.
- MARIA (STF):[70] The Brazilian Supreme Court (STF) launched MARIA (Módulo de Apoio para Redação com Inteligência Artificial),

67. *ASSIS*, *supra* note 26.

68. Nathália Valente, *Nova Versão do Chat-JT Conta Com Integração ao PJe* [*New Version of Chat–JT Features PJe Integration*], CONSELHO SUPERIOR DA JUSTIÇA DO TRABALHO [SUPERIOR COUNCIL OF LABOR JUSTICE] (May 6, 2025), https://www.csjt.jus.br/web/csjt/-/nova-vers%C3%A3o-do-chat-jt-conta-com-integra%C3%A7%C3%A3o-ao-pje.

69. STJ Lança Novo Motor de Inteligência Artificial Generativa Para Aumentar Eficiência Na Produção de Decisões [*STJ Launches New Generative Artificial Intelligence Engine to Increase Efficiency in Decision–Making*], SUPERIOR TRIBUNAL DE JUSTIÇA [SUPERIOR COURT OF JUSTICE] (Feb. 12, 2025), https://www.stj.jus.br/sites/portalp/Paginas/Comunicacao/Noticias/2025/11022025-STJ-lanca-novo-motor-de-inteligencia-artificial-generativa-para-aumentar-eficiencia-na-producao-de-decisoes.aspx.

70. Jorge Macedo, STF Lança MARIA, Ferramenta de Inteligência Artificial Que Dará Mais Agilidade Aos Serviços do Tribunal [*STF Launches MARIA, An Artificial Intelligence Tool That Will Streamline Court Services*], SUPREMO TRIBUNAL FEDERAL [SUPREME COURT] (Dec. 16,

> a tool integrated into the STF-Digital system. MARIA automatically generates draft summaries and headnotes (ementas) for judicial opinions, which can then be edited and validated directly by the justices.

These institutional initiatives reflect a broader shift toward the systematization of GenAI tools within judicial governance, combining technological innovation with constitutional and procedural safeguards. Rather than displacing judicial authority, these tools are designed to augment legal analysis, streamline internal workflows, and standardize outputs—particularly in high-volume courts.

b) Court Management

Within Brazil's higher courts, predictive AI has been strategically applied to optimize internal workflows and support judicial decision-making. Two systems in particular stand out for their scope and impact: VICTOR, developed by the Supreme Federal Court, and Athos, created by the Superior Court of Justice. While both systems leverage natural language processing and machine learning, they operate in distinct institutional contexts and pursue different goals—VICTOR focuses on constitutional admissibility, whereas Athos targets the identification and management of repetitive legal issues.

VICTOR, introduced in 2019, is one of the earliest and most emblematic examples of AI applied to court management. Developed by the Supreme Court, it was designed to assist in identifying cases involving general repercussion—a constitutional admissibility requirement for extraordinary appeals. Trained on 28 established general repercussion themes with sufficient jurisprudential data, the system could flag whether a given case potentially involved one of these themes. When uncertain, the system would return an "inconclusive" result. Regardless of the output, the final determination remained subject to human review by a civil servant.[71]

To perform this task, VICTOR applies optical character recognition (OCR) to all incoming cases, enabling the system to identify and reorganize key documents within each file. This significantly improves the structure and accessibility of case records in the STF's digital system.

Another example of AI applied to court management is Athos, a system developed by the Superior Court of Justice (STJ), Brazil's highest court for nonconstitutional matters. Athos was created to help the court identify and

---

2024), https://noticias.stf.jus.br/postsnoticias/stf-lanca-maria-ferramenta-de-inteligencia-artificial-que-dara-mais-agilidade-aos-servicos-do-tribunal/.

71. Fabiano Hartmann & Debora Bonat, *Machine Learning and the General Repercussion on Brazilian Supreme Court: Applying the Victor Robot to Legal Texts* 7–8 (2020), https://ceur-ws.org/Vol-2632/MIREL-19_paper_5.pdf.

manage repetitive legal issues, especially those eligible for resolution through Brazil's system of binding precedents. It uses natural language processing and semantic clustering to group appeals that involve similar legal questions. Its main features include similarity search, keyword-based search, monitoring of legal controversies, case clustering, and support for admissibility analysis. These tools improve workflow efficiency and promote consistency in judicial decisions.

From 2020 to 2021, Athos was used in 40% of all new legal controversies formally recognized by STJ. In the Brazilian legal system, a "legal controversy" refers to a recurring legal question that appears in multiple cases and may be resolved through a single precedent-setting decision. During this period, the number of internal requests to use Athos grew by 211%, reflecting increased reliance on the system by court staff and legal analysts.[72]

Athos has also contributed to a measurable reduction in caseloads. By mid-2021, more than 350,000 cases were resolved in lower courts—through settlements, withdrawals, or decisions not to appeal—without being sent to the STJ. In cases involving the federal government, the number of special appeals filed by the Office of the Attorney General (AGU) fell by 11.2% compared to the same period the year before. The rate of unfavorable decisions in these cases decreased by 14.15%, and over 1,400 appeals were voluntarily withdrawn from the STJ.[73]

c) Interfacing with the Public

Despite the emphasis on backend efficiency, several federal courts across Brazil have independently developed AI-powered virtual assistants to enhance the experience of those seeking justice. From Espírito Santo to Rio de Janeiro, tools such as Fale com a Ju, Judi, and Justa have emerged as emblematic examples of this citizen-facing innovation. Though developed in distinct institutional contexts and serving different immediate needs, these initiatives share a common aim: bridging the gap between the judiciary and the public by offering accessible, responsive, and human-centered digital interfaces.

72. *Ascom, Sistema do STJ Que Automatiza Fluxo de Processos Começa a Operar No TJMA* [*STJ System That Automates Process Flow Begins Operating at TJMA*], PORTAL DO PODER JUDICIÁRIO: DO ESTADO DO MARANHÃO [JUDICIAL PORTAL: STATE OF MARANHÃO] (June 27, 2023), https://www.tjma.jus.br/midia/portal/noticia/510543/sistema-do-stj-que-automatiza-fluxo-de-processos-comeca-a-operar-no-tjma.

73. Guilherme Silva Figueiredo, *Projeto Athos: Um Estudo de Caso Sobre a Inserção do Superior Tribunal de Justiça Na Era da Inteligência Artificial* [*Project Athos: A Case Study on the Superior Court of Justice's Integration Into the Age of Artificial Intelligence*], UNIVERSIDADE DE BRASÍLIA [UNIVERSITY OF BRASILIA] 102 (2022), https://www.cnj.jus.br/wp-content/uploads/2022/11/projeto-athos.pdf.

Launched during the COVID-19 pandemic, Fale com a Ju ("Talk to Ju") was created by the Federal Court of Espírito Santo to respond to the surge in litigation related to emergency aid benefits.[74] Deployed via WhatsApp, the assistant provided clear information on eligibility criteria, addressing widespread documentation issues that had led to mass denials. By offering automated responses, it helped reduce unnecessary claims and guided legitimate cases more efficiently into the judicial system.

Similarly, the Federal Court in Rio de Janeiro (JFRJ) launched "Judi," a virtual assistant accessible through Telegram that redirects users to official information on the court's website, enhancing accessibility and user autonomy.[75]

In the same spirit, the 27th Federal Court of Rio de Janeiro (JFRJ) implemented *Justa*, an AI-powered virtual assistant designed to improve interaction with court users. Accessible at all hours via WhatsApp and Instagram, *Justa* provides case updates, clarifies procedural steps, and responds to common inquiries, particularly benefiting self-represented litigants and those facing structural barriers to accessing the justice system.[76]

Justa is part of a broader project titled "VIC–Vara Integrada ao Cidadão"[77] ("Court Integrated with the Citizen"), led by the presiding judge, Hon. Geraldine de Castro. The initiative seeks to align judicial services with principles of inclusivity and institutional transparency, in line with the United Nations Sustainable Development Goals. Rather than focusing solely on automation, the project emphasizes enhancing communication channels between the court and the broader public.[78]

---

74. InovarES - Laboratório de Inovação da Justiça Federal do Espírito Santo [*InovarES–Innovation Laboratory of the Federal Court of Espírito Santo*], RENOVAJUD, https://renovajud.cnj.jus.br/laboratorios-publico?laboratorio=12&utm; *JFES Lança Atendimento por Whatsapp Com a Utilização de Chatbot* [*JFES Launches WhatsApp Support Using Chatbot*], ASSOCIAÇÃO DOS JUÍZES FEDERAIS DO BRASIL [BRAZILIAN FEDERAL JUDGES ASSOCIATION] (May 27, 2020), https://www.ajufe.org.br/imprensa/noticias-do-judiciario/13923-jfes-lanca-atendimento-por-whatsapp-com-a-utilizacao-de-chatbot.

75. *Atendimento Virtual* [*Virtual Assistance*], JUSTIÇA FEDERAL 2ª REGIÃO [FEDERAL COURT 2ND REGION] (Feb. 15, 2024), https://www.trf2.jus.br/jfrj/duvidas-frequentes/atendimento-virtual.

76. *Justa - Assistente Virtual* [*Justa–Virtual Assistant*], JUSTIÇA FEDERAL 2ª REGIÃO [FEDERAL COURT 2ND REGION] (Sep. 20, 2024), https://www.trf2.jus.br/juizo/jfrj/27vf/justa-assistente-virtual.

77. *Vara Integrada ao Cidadão–VIC* [*Integrated Citizen Court–VIC*], JUSTIÇA FEDERAL 2ª REGIÃO [FEDERAL COURT 2ND REGION] (Sep. 21, 2024), https://www.trf2.jus.br/jfrj/artigo/27vf/vara-integrada-ao-cidadao-vic.

78. *JF 2ª Região, Por Meio de Iniciativa Inovadora da 27ª VF/RJ, Está Incluída No Portal de Boas Práticas do Judiciário/CNJ* [*Federal Court 2nd Region, Through an Innovative Initiative of the 27th Federal Court of Rio de Janeiro, Is Included in the Judiciary/CNJ Best Practices Portal*], JUSTIÇA FEDERAL 2ª REGIÃO [FEDERAL COURT 2ND REGION] (Aug. 19, 2024), https://www.trf2.jus.br/trf2/

Innovation is further structured through *Lab27*, an internal unit dedicated to experimental solutions developed collaboratively by court staff and trainees.[79] Among its outcomes are simplified digital forms for urgent procedural requests and redesigned citation templates using plain language. These developments reflect the capacity for administrative innovation within Brazil's judiciary, where trial-level judges exercise broad discretion in managing court operations and public service delivery.

B. CHINA AI-ENHANCED COURTS AT SCALE

As in Brazil, the judiciary in China faces immense pressures of scale, and therefore, immense incentives and opportunities to expand the reach of the courts through AI. Rising demand and limited human capacity, the country's broader digital and artificial intelligence ambitions,[80] and investments in data and technical infrastructure have led to China's adoption of AI to the point where it has been identified as "probably the most advanced and prolific judicial user of AI."[81]

China's vast, stratified, and heterogeneous judicial system includes elite, tech-enabled tribunals in urban centers and severely under-resourced county courts serving remote and rural populations,[82] as represented by the Supreme People's Court (SPC), 31 High People's Courts, Intermediate People's Courts, and over 3,000 Basic People's Courts.[83] China has an estimated 650,000 lawyers for a population of more than 1.4 billion, or about one lawyer for about 4,000

noticia/2024/jf-2a-regiao-por-meio-de-iniciativa-inovadora-da-27a-vfrj-esta-incluida-no-portal.

79. *Lab27*, JUSTIÇA FEDERAL 2ª REGIÃO [FEDERAL COURT 2ND REGION] (Sep. 30, 2024), https://www.trf2.jus.br/jfrj/artigo/27vf/lab27.

80. *See* Rachel E. Stern, Ben L. Liebman, Margaret E. Roberts & Alice Z. Wang, *Automating Fairness? Artificial Intelligence in the Chinese Courts*, 59 COLUM J. TRANSNAT'L L. 515, 530 (2021) ("China's push for [artificial intelligence] is an important part of the country's strategic response to slowing economic growth," on the one hand, and "motivated by a pervasive belief in nationalist vindication through technological innovation" on the other. "Viewed through this lens, the courts' strides toward algorithmic analytics contribute to the 'first in the world' narrative of technological success poised to become a prominent part of the Party's twenty-first century legitimacy strategy."); *see also* Zhiyuan Guo & Jiajia Yang, *The Application of Artificial Intelligence in China's Criminal Justice System*, 6 LEGAL ISSUES DIGIT. AGE 83, 83–104 (2025) (describing assorted government efforts to build a "Digital China").

81. Gary E. Marchant, *AI in Robes: Courts, Judges, and Artificial Intelligence*, 50 OHIO N.U. L. REV. 473, 486 (2024).

82. *China*, JUDICIARIES WORLDWIDE: A RES. ON COMPAR. JUD. PRAC., https://judiciariesworldwide.fjc.gov/country-profile/china#:~:text=China%20has%20a%20unified%20court,cases%20from%20their%20territorial%20designations; Xin Dai, *Who Wants a Robo-Lawyer Now?: On AI Chatbots in China's Public Legal Services Sector*, 26 YALE J.L. & TECH. 527, 535 (2024).

83. *China*, JUDICIARIES WORLDWIDE, *supra* note 82.

people.[84] This gap is most pronounced in rural areas as professionals trained in legal services have historically gravitated toward developed urban areas.[85] Furthermore, since 1978, there has been roughly a 30-fold increase in cases but only a 3-fold increase in judges, with growth in the judiciary limited by the need to increase the legitimacy and professionalism of the judiciary.[86]

The mismatch between the demand for legal services and professional supply has led China's Ministry of Justice (MOJ) to pursue reform initiatives focused on establishing a "public legal services system" that "covers all urban and rural residents," in order to "ensure that people receive timely and effective legal help."[87] Alongside pro bono representation in litigation, the MOJ seeks to provide public legal services in the form of providing basic legal information and answering routine questions.[88] In later years, this goal was transformed into numerical targets with the objective of having onsite legal advisors for the country's more than 690,000 rural villages and nearly 120,000 urban residential communities.[89] Achieving this goal with human professionals alone sits somewhere between extremely expensive and unlikely to nearly impossible, presenting the opportunity for AI tools to offer accessible, consistent legal services at scale to populations who would otherwise receive no services.

At the same time, several factors have contributed to the embrace of legal tech across the judicial system in China. Government initiatives have led to the development of "Smart Courts" that make use of technology and data, online trial services, and AI-powered services integrated with ubiquitous platforms like WeChat.[90] The state has invested in streamlining judicial processes through automation and digitization,[91] and it made no secret of its desire to lead the world in artificial intelligence as well as to elevate the role, presence, and the image of the judiciary. This top-down support has given both the judiciary and legal practitioners the confidence and incentive to introduce AI-based tools.[92]

84. Dai, *supra* note 82, at 536.

85. *Id.* at 535–36.

86. Nyu Wang & Michael Yuan Tian, *"Intelligent Justice": Human–Centered Considerations in China's Legal AI Transformation*, 3 AI ETHICS 349, 350 (2023).

87. Dai, *supra* note 82, at 535.

88. *Id.*

89. *Id.* at 536.

90. Sebastian Ko, *5 Factors Driving the Chinese Lawtech Boom*, WORLD ECON. F. (Apr. 1, 2019), https://www.weforum.org/stories/2019/04/5-factors-driving-the-chinese-lawtech-boom/.

91. *See, e.g.*, Mimi Zou, "*Smart Courts" in China and the Future of Personal Injury Litigation*, J. PERS. INJ. L. 1, 2 (2020) (describing the development of open online judicial information platforms for the publication of all judicial documents produced by Chinese courts).

92. *Id.*

Structural, political, and cultural factors have also played an important role. Automated systems make it easier to monitor judges and standardize approaches. China's legal system is still young compared to those of other major economies, supporting experimentation.[93] The growth in public demand for legal services has been driven by new laws, rising disputes, and expanding awareness of legal rights. A smartphone-centric culture has further pushed the delivery of legal services into mobile formats. With a limited legacy infrastructure to disrupt, Chinese legal professionals have the potential to leapfrog conventional legal systems. Together, these factors have contributed to a government-supported ecosystem for judicial AI that has prioritized accessibility, scale, and efficiency, though potentially over other important considerations like fairness, accountability, and judicial authority.[94]

1. Core Judicial Functions

China has integrated AI into its core judicial functions for some time, in national, provincial, and municipal courts. Below, we highlight the Same Type Case Reference system (STCR) and FaXin at the national level, as well as local examples, particularly Shanghai's "System 206" and Hangzhou's Xiao Zhi 3.0. Much of the groundwork for the use of AI in the Chinese legal system has been laid by the country's national "Big Data" and related strategies.[95] For example, starting in 2014, the Supreme People's Court (SPC) mandated the public disclosure of judicial decisions, which were hosted on a centralized website called "China Judgments Online."[96] Although compliance with this mandate and comprehensive accountability remains a work in progress,[97] the platform hosted over 160 million documents at the time of this writing.[98]

93. Jinting Deng, *Should the Common Law System Welcome Artificial Intelligence: A Case Study of China's Same-Type Case Reference System*, 3 GEO. L. TECH. REV. 223 (2019) (describing how modern Chinese caselaw began in the 1980s).

94. *See, e.g.*, Nyu Wang et al., *supra* note 86, at 351 (describing challenges to the implementation of AI in the Chinese courts as including unproven technologies, uneven availability of case data, and a lack of accountability); *see also* Straton Papagianneas & Nino Junius, *Fairness and Justice Through Automation in China's Smart Courts*, 51 COMPUT. L. & SEC. REV. 1, 2–3 (2023) (surveying critical analyses of China's smart court initiatives and identifying due process, transparency, judicial independence, and fairness concerns).

95. *See* China's State Council, Promotion of Big Data Development Action Outlines in 2015, the Industry Ministry 2016–2020 Plan for the Development of the Big Data Industry in 2016, and subsequent initiatives.

96. *See* Stern et al., *supra* note 80, at 522.

97. *See id.* at 533–37.

98. CHINA JUDGEMENTS ONLINE, https://wenshu.court.gov.cn/ (last visited Feb. 2, 2026) (reporting 163M documents).

Commitments to digital services and big data have made possible the introduction of the STCR in 2015, a series of systems for promoting judicial uniformity by making prior cases binding on subsequent courts.[99] The infrastructure of the system includes a national database of continuously updated judicial decisions, software for searching and locating comparable cases among millions of judgments, and accountability rules that require cases to be decided in a way consistent with those previously decided.[100] The system allows judges to upload complaints or hearing records and retrieve comparable precedent cases.[101] The system can suggest outcomes, calculate expected sentences, and summarize typical remedies—effectively importing a form of precedent-based reasoning into a civil law framework. Although consistency in judicial outcomes is traditionally considered helpful in upholding democratic values, part of the purpose of STCR was and continues to be to strengthen supervision over ordinary judges and restrict judicial discretion. Most of the focus in this area has been on the vast amount of data that has been publicly released, without a full accounting for the data and cases that are missing/not public. Since the database of previous decisions informs this "precedent-based" reasoning, understanding and evaluating how complete the record is should remain a priority.[102] Although implementations of STCR vary by province, they rely heavily on computer and AI logics to provide case recommendations (through a "same-type" computer program), statistical analysis of prior analogous cases (through a "prior-case" analysis program), and draft judgments (through a "judgment-generating" program).[103]

Another major AI initiative has been Shanghai's "System 206," officially known as the Trial-centered Litigation Reform Software.[104] Its development began in 2017 under the umbrella of China's broader "Smart Justice" initiative, launched by the Supreme People's Court in 2016 to modernize the judiciary through digital tools, big data, and artificial intelligence.[105] Jointly created by iFlytech and Shanghai's judicial, procuratorial, and public security agencies, System 206 is an AI judicial assistant that can support various phases of the criminal process. It helps judges with fact finding, authenticating evidence, and

99. Deng, *supra* note 93, at 225.

100. *Id.* at 237.

101. *Id.* at 226.

102. *See* Stern et al., *supra* note 80, at 534.

103. Deng, *supra* note 93, at 252.

104. *See* YADONG CUI, ARTIFICIAL INTELLIGENCE AND JUDICIAL MODERNIZATION ix–xi (Cao Yan & Liu Yan trans., Springer 2020).

105. Wanqiang Wu & Xifen Lin, *Access to Technology, Access to Justice: China's Artificial Intelligence Application in Criminal Proceedings*, 81 INT'L J.L., CRIME & JUST. 1, 1 (2025).

improving consistency in criminal cases.[106] The 206 System can accept verbal commands to display relevant information on digital screens.[107] It is also able to transcribe speech throughout hearings and identify speakers according to their roles as judges, prosecutors (referred to in China as "procurators"), and defendants.[108] It was the first court to experiment with AI in adjudication.[109] But while its primary goals are to standardize evidence collection, reduce discretionary inconsistencies, studies of its operations suggest that the system has not necessarily lived up to its promise, its benefits have been unevenly realized, and its usage has concentrated in simple, high-volume cases.[110]

Another early and highly publicized system, in use in Hangzhou since 2019, is Xiao Zhi 3.0, or "Little Wisdom."[111] Although the system was originally just used for simple financial adjudications, its abilities have grown over time.[112] Today, it analyzes filings, summarizes disputes, evaluates evidence, and drafts judicial documents. In one widely publicized case, Xiao Zhi was used to hear and resolve a case, start to finish, in under 30 minutes.[113] Through remote proceedings, a "Robojudge," not only assists judges, but can also decide ecommerce, product liability and copyright disputes.[114] In some contexts, human judges that disagree with AI rulings are required to submit written explanations,[115] though it is unclear the degree to which judicial decisions, overall, are AI-assisted. Additionally, it is unclear whether litigants have the opportunity to opt-out of or object to the use of AI systems in their matters.[116]

These initiatives are poised to continue to proliferate. In 2024, China's Supreme People's Court introduced the FaXin foundation model system, a national large AI model designed to enhance the efficiency and application of

106. Jiang Wei, *China Uses AI Assistive Tech on Court Trial for First Time*, CHINA DAILY (Jan. 24, 2019), https://www.chinadaily.com.cn/a/201901/24/WS5c4959f9a3106c65c34e64ea.html.

107. Liang Chenyu, *Shanghai Court Adopts New AI Assistant*, SIXTH TONE (Jan. 25, 2019), https://www.sixthtone.com/news/1003496.

108. *Id.*

109. *Id.*

110. Wu et al., *supra* note 105, at 13; *see also* Stern et al., *supra* note 80, at 543.

111. Alena Zhabina, *How China's AI Is Automating the Legal System*, DEUTSCHE WELLE (Jan. 20, 2023), https://www.dw.com/en/how-chinas-ai-is-automating-the-legal-system/a-64465988.

112. *How Is China Using AI?*, APAC INSIDER (May 18, 2023), https://apacinsider.digital/how-is-china-using-ai/.

113. Zhabina, *supra* note 111.

114. Hadar Y. Jabotinsky & Michal Lavi, *AI in the Courtroom: The Boundaries of RoboLawyers and RoboJudges*, 35 FORDHAM INTELL. PROP., MEDIA & ENT. L.J. 286, 291 (2025).

115. *Id.* at 385.

116. Zhiyu Li, *AI and Human Judges in Chinese Courts* 7 (Jan. 10, 2025), http://dx.doi.org/10.2139/ssrn.5235753 (on file with the Berkeley Technology Law Journal).

AI technologies within the legal field, such as improving judicial paperwork processes.[117] The model's database contains 320 million entries, including legal documents, court judgments, cases, articles, and related materials, amounting to 3.67 trillion Chinese characters on multiple legal data platforms. Like STCR, FaXin seeks to strengthen supervision over ordinary judges and restrict judicial discretion.[118] The system issues an alert for supervisory review whenever a judge's decision differs from the AI recommendation.[119] Every action is timestamped and tied to the judge's final performance review, incentivizing judges to align with AI recommendations.[120] Finally, litigants see the final judgment but have no insight into the data or algorithms that shaped the decision.[121] Although this system arguably reduces idiosyncratic bias, it entrenches systemic conformity bias, particularly since the AI-influenced decisions of today become tomorrow's training data.[122]

2. Court Management

In addition to assisting judges, AI is also being deployed in China to support court operations, improving the administrative efficiency of judicial institutions. The Xiao Zhi 3.0 system supports not only case analysis but also logistical tasks like scheduling hearings and announcing court procedures. More broadly, the FaXin foundation model helps automate judicial paperwork and assists in training and standardization efforts by embedding best practices into its recommendations. Systems like STCR, while designed with jurisprudential aims, also serve a managerial role by ensuring compliance with centrally approved legal interpretations and reducing inconsistencies among lower court decisions.[123]

Case management is another realm to which data and AI have been applied. Various Chinese provinces have developed AI-based case management systems to meet different purposes. For example, the Zhejiang People's Procuratorate worked with Alibaba Cloud to co-develop a big data platform that visualizes case data in dynamic charts to support prosecutorial decisions.[124] In Beijing, a similar platform integrates litigation-stage data, enabling fast access to legal documents. Other provinces—like Guizhou,

117. Huaxia, *China Unveils AI Model to Facilitate Judicial Work*, XINHUANET (Nov. 15, 2024), https://english.news.cn/20241115/20c393c7d1a4441cad9581125cad4561/c.html.

118. Ernest Lim & Ilya Akdemir, *Same Words, Different Worlds: The Illusion of Shared Judicial AI Principles* (forthcoming) (manuscript at 22) (on file with authors).

119. *Id.*

120. *Id.*

121. *Id.* at 22–24.

122. *Id.* at 34.

123. *See* Deng, *supra* note 93, at 233–36.

124. Guo et al., *supra* note 80, at 85.

Hainan, Yunnan, Jiangsu, and Guangdong—are also developing their own AI-based systems, reflecting the broad integration of AI into case handling and court operations in criminal justice contexts.[125] "Smart Court" systems in, for example Hainan, Guizhou, Yunnan and Guangzhou provinces, have also applied AI to carry out a variety of functions including "litigation service reception, case file transfer, pretrial meetings, trial recording" and a variety of evidence-related tasks.[126]

3. Interfacing with the Public

China's Smart Courts initiative has also included a variety of mechanisms for making the court more accessible to ordinary users. This is most visible in the Internet Courts, first launched in Hangzhou in 2017 and later expanded to Beijing and Guangzhou.[127] Designed to handle online disputes efficiently,[128] these courts allow litigants to file, mediate, and resolve cases entirely online. Parties can access court services through a WeChat-based "mobile micro court" app that uses facial recognition to authenticate identities and allows for communication with judges via text or audio, uploading of evidence, and electronic signing of documents.[129] Pretrial mediation is integrated into the platform and, if successful, results in a binding agreement without the need for formal litigation. The Internet Courts also incorporate blockchain technology to store and authenticate digital evidence, a practice the Supreme People's Court has endorsed as legally valid when certain conditions are met. These courts function not only as a response to practical burdens—like geographic dispersion and overloaded dockets—but also as a testbed for more radical digital transformations in adjudication.

To bring legal information to the parties, the Beijing Internet Court has touted the use of digital assistants, even in hologram form, for providing basic guidance on laws, court procedures, and whether a court has jurisdiction or if alternative dispute resolution is more appropriate. Courts in a number of provincial level regions, including Beijing, Shanghai, and Guangdong, have also introduced AI-powered service robots in court halls. These bots help users navigate litigation manuals, judicial procedures, and offer information

125. *Id.*

126. *Id.* at 91.

127. Zou, *supra* note 91, at 4.

128. It has been reported that in Beijing, "the average duration of a case is 40 days; the average dispositive hearing lasts 37 minutes; almost 80 per cent of the litigants before the Chinese Internet courts are individuals, and 20 per cent corporate entities; and 98 per cent of the rulings have been accepted without appeal." Tara Vasdani, *Robot Justice: China's Use of Internet Courts*, LAW360 CANADA, (Feb. 5, 2020), https://www.law360.ca/ca/articles/1750396/robot-justice-china-s-use-of-internet-courts.

129. *See* Guo et al., *supra* note 80, at 85; *see also* Zou, *supra* note 91, at 4.

about specific judges and clerks. Some advanced models can even predict likely outcomes for a party before a case is formally filed, illustrating the growing sophistication of AI in legal access tools.[130]

Additionally, the FaXin model is accessible to the public, offering litigants AI-powered legal advice tailored to their specific situations. This marks a shift from purely internal judicial AI to tools that aim to close the information and guidance gaps experienced by unrepresented users.

C. THE UNITED STATES: JUDICIAL EXPERIMENTATION AND INNOVATION

In contrast to China and Brazil, in the United States there has been less uptake of AI systems built specifically for judges, and more piloting and experimentation across a variety of use cases. This is due in no small part to the decentralized nature of the American judiciary; along with one Supreme Court, 13 appellate courts, 94 trial court districts, and a handful of specialty courts at the federal level,[131] there are approximately 15,000 to 17,000 different state and municipal courts,[132] and a concomitant lack of uniformity with respect to procurement and use policies for adopting Al.[133] In the one realm in which there has been relatively high uptake, the adoption of algorithms in risk assessment contexts, there has also been vigorous debate and contestation, due to concerns about algorithmic bias, opacity, and the risk of due process violations.[134] A vibrant legal tech sector[135] and the integration of AI into

130. Zou, *supra* note 91, at 3.

131. *Court Role and Structure*, U.S. COURTS, https://www.uscourts.gov/about-federal-courts/court-role-and-structure.

132. Cary Coglianese & Lavi M. Ben Dor, *AI in Adjudication and Administration*, 86 BROOK. L. REV. 791, 794 (2021).

133. *See* Simshaw, *supra* note 6, at 806.

134. *See* Coglianese et al., *supra* note 132, at 805–11 (describing challenges to risk assessment tools). An evaluation by Jennifer Doleac and Megan Stevenson found that the use of risk assessments in sentencing contexts did not translate into gains in public safety or reductions in rates of incarceration, in part because of the ways in which judges deviated from tool recommendations. *See generally* Megan T. Stevenson & Jennifer L. Doleac, *Algorithm Risk Assessment in the Hands of Humans*, 16 ECON. POL'Y 382 (2024). Judges were also more likely to follow leniency recommendations for White rather than for Black defendants. *Id.* A recent analysis of 27,357 sentencing cases of drug offenses in Virginia between 2013 and 2022 concluded that a risk assessment tool's recommendations impacted judicial fairness in opposite directions, alleviating gender-based disparity in favor of females, but triggering racial bias favoring White over Black offenders. Yi-Jen (Ian) Ho, Wael Jabr & Yifan Zhang, *AI Enforcement: Examining the Impact of AI on Judicial Fairness and Public Safety* 4–6 (2024) (manuscript), https://ssrn.com/abstract=4533047.

135. As measured, for example, by legal tech funding. *See, e.g.*, Steven Lerner, *Legal Tech Sees 80% Funding Surge Amid AI Boom*, LAW360 (Apr. 7, 2025), https://www.law360.com/pulse/articles/2321847/legal-tech-sees-80-funding-surge-amid-ai-boom (describing the United States as the top country for receiving legal tech investment over the studied period).

mainstream legal research platforms have bolstered the uptake of AI by practicing lawyers.[136] But litigant use of AI has not necessarily built confidence in the capacity of generative artificial intelligence to support high quality work product, with numerous high-profile cases of lawyers filing briefs that contain hallucinated citations,[137] at least one of which has, unfortunately, made it into a court decision.[138] These incidents have, in turn, placed emphasis on the *responsive* role of U.S. courts to litigant use of AI, whether through the development of standing orders governing lawyers' uses of generative tools[139] or the recent proposed introduction of a new rule, Federal Rule of Evidence 707, to regulate the authentication and admission of evidence that is or is suspected to be machine-generated.[140]

Nevertheless, the conditions for the increased, *proactive* adoption of AI by judges and courts are also present. Electronic filings and digitalization, as well as the publication of court proceedings mean that massive troves of court records, pleadings, and decisions are available for training machines for both "LegalTech" and "JudgeTech" applications.[141] There is a yawning access to

136. *See, e.g.*, *LexisNexis Announces Launch of Lexis+ AI Commercial Preview, Most Comprehensive Global Legal Generative AI Platform*, LEXISNEXIS (May 4, 2023), https://www.lexisnexis.com/community/pressroom/b/news/posts/lexisnexis-announces-launch-of-lexis-ai-commercial-preview-most-comprehensive-global-legal-generative-ai-platform (discussing how LexisNexis announced the launch of its Lexis+ AI platform in May 2023); *see also Thomson Reuters Debuts Westlaw Precision*, PR NEWSWIRE (Sep. 14, 2022), https://www.prnewswire.com/news-releases/thomson-reuters-debuts-westlaw-precision-301624347.html (discussing how Westlaw Precision, which also incorporates generative AI, was introduced the year before).

137. As of January 31, 2026, Damien Charlotin has collected over 600 legal decisions in the U.S. in which generative AI produced hallucinated content, including both fake citations and other types of content. *See AI Hallucination Cases*, DAMIEN CHARLOTIN, https://www.damiencharlotin.com/hallucinations/.

138. Shahid v. Esaam, 376 Ga. App. 145, 145 (2025) (describing the trial court's use of "an order that relied upon non-existent case law").

139. Tracked, for example, by the law firm Ropes and Gray; by July 2025, its database of standing orders, local rules, or judicial decisions included 396 entries as of January 31, 2026. Amy Jane Longo, Shannon Capone Kirk & Isaac Sommers, *Standing Court Order Tracker*, ROPES & GRAY, https://www.ropesgray.com/en/sites/artificial-intelligence-court-order-tracker.

140. *See* Shane Ramsey, *Safeguarding the Courtroom from AI-Generated Evidence: Federal Rule of Evidence 707 Approved by Judicial Conference*, JD SUPRA (June 13, 2025), https://www.jdsupra.com/legalnews/safeguarding-the-courtroom-from-ai-1931550/ (describing that the new rule essentially requires machine-generated evidence to meet the same standard of admissibility as traditional expert evidence).

141. However, comprehensive access to federal court records generally requires a subscription to a commercial database or surmounting a government paywall, PACER, a problem that notable court record liberation efforts like the SCALES project are trying to address. David L. Schwartz, Kat M. Albrecht, Adam R. Pah, Christopher A. Cotropia, Amy Kristin Sanders, Sarath Sanga, Charlotte S. Alexander, Luís A.N. Amaral, Zachary D. Clopton, Anne M. Tucker, Thomas W. Gaylord, Scott G. Daniel & Nathan Dahlberg, *The SCALES*

justice gap—according to the Legal Services Corporation, "[l]ow-income Americans do not get any or enough legal help for 92% of their substantial civil legal problems."[142] Parties without legal training frequently are in direct contact with the court. As has been observed, the typical American civil trial court is "lawyerless"—in contrast to the situation within federal courts, "more than 75% of state court claims involve at least one party without legal representation."[143] Decentralization also supports experimentation as technology can be initially tried on a small scale without requiring systems-wide change.[144] In addition, as litigants increasingly turn to AI tools to carry out legal research, generate pleadings, and support the production of evidence,[145] it is inevitable that courts, too, will increasingly incorporate AI tools into their work flows. Indeed, over 50 courts at the time of this writing are reportedly actively testing AI tools for legal research, document review, and case management.[146]

Guidance issued by members of the American Bar Association (ABA) Task Force on Law and Artificial Intelligence Working Group on AI and the Courts has sanctioned doing so. While not formally endorsed by the ABA, the document nonetheless represents a thoughtful attempt to outline the judicial tasks to which the application of AI is recommended.[147] These include, in the realm of core judicial functions:

- Conducting legal research, when appropriately trained;
- Drafting routine orders;
- Searching and summarizing discovery;
- Creating event timelines; and

Project: Making Federal Court Records Free, 119 NW. U. L. REV. 23, 32–37 (2024) (describing the history of access to court records, and the SCALE project). There is no unified interface for accessing state court records.

142. LEGAL SERVS. CORP., THE JUSTICE CAP: THE UNMET CIVIL LEGAL NEEDS OF LOW-INCOME AMERICANS 7 (2022), https://lsc-live.app.box.com/s/xl2v2uraiotbbzrhuwtjlgi0emp3myz1.

143. Diego A. Zambrano, *Missing Discovery in Lawyerless Courts*, 122 COLUM. L.J. 1423, 1425 (2022).

144. Madison Alder, *U.S. Court System Eyeing AI Use Cases for Access to Justice, Cost Savings*, FEDSCOOP (May 6, 2025), https://fedscoop.com/u-s-court-system-eyeing-ai-use-cases-for-access-to-justice-cost-savings/ (describing the federated nature of the federal judiciary as providing opportunities for individual circuits and courts to experiment, learn, and teach).

145. *See* Wilf-Townsend et al., *supra* note 3, at 2–6.

146. NAT'L CTR. FOR STATE CTS., *Guidance for Implementing AI in Courts*, https://www.ncsc.org/resources-courts/guidance-implementing-ai-courts (reporting that 50+ courts are "actively testing AI tools for legal research, document review, and case management").

147. Hon. Herbert B. Dixon Jr, Hon. Allison H. Goddard, Maura R. Grossman, Hon. Xavier Rodriguez, Hon. Scott U. Schlegel & Hon. Samuel A. Thumma., *Navigating AI in the Judiciary: New Guidelines for Judges and Their Chambers*, 26 SEDONA CONF. J. 1 (2025).

- Evaluating submissions by the parties for legal sufficiency.

In the realm of court management:

- Generating court notices and communications;
- Court scheduling and calendar management;
- Translation of foreign language documents and transcription of court proceedings;
- Analysis of court operations; and
- Document organization and management.[148]

In the realm of public facing AI, the guidance also encourages the use of AI to enhance court accessibility services and assist pro se litigants.[149]

This work builds on the longstanding coordination efforts of the National Center for State Courts (NCSC) among state courts seeking to adopt AI. In 2020, the NCSC published "Introduction to AI for Courts," a short document that described some early examples of the use of AI by U.S. courts.[150] Following the introduction of ChatGPT, the NCSC launched, in coordination with the Conference of Chief Justices (CCJ) and the Conference of State Court Administrators (COSCA), the "AI Rapid Response Team" (RRT) to help respond to the new set of issues posed by generative AI.[151] In addition to a series of short "interim" guidance documents on various topics,[152] it released, in 2024, "Artificial Intelligence: Guidance for Use of AI and Generative AI in Courts."[153] Part primer and part guidance document, the RRT advises courts that are exploring the use of generative AI to take several steps: develop an internal use policy, start with a few low risk tasks, and use a "human-in-the-loop" or "human-on-the-loop" approach.[154]

148. *Id.* at 6–7.

149. *Id.*

150. JOINT TECH. COMM., *JTC Resource Bulletin: Introduction to AI for Courts* 2, 7 (Mar. 27, 2020), Nat'l Ctr. for State Cts., https://ncsc.contentdm.oclc.org/digital/collection/tech/id/930 (describing, for example, the use of facial recognition-based sign-on by Marion County judges, a New Jersey court chatbot called JIA); *see also* Marchant, *supra* note 81, at 481.

151. *See* Marchant, *supra* note 81, at 481–82.

152. *See, e.g.*, NAT'L CTR. FOR STATE CTS., AI AND THE COURTS: GETTING STARTED (Mar. 2024), https://www.ncsc.org/sites/default/files/media/document/RRT-AI-getting-started-march-2024.pdf.

153. *See generally* NAT'L CTR. FOR STATE CTS., ARTIFICIAL INTELLIGENCE: GUIDANCE FOR USE OF AI AND GENERATIVE AI IN COURTS (Aug. 7, 2024), https://nationalcenterforstatecourts.app.box.com/s/65mh1qmyx9ap469kjj386vhk0vxtpral.

154. *Id.* at 13–16.

In Fall 2024, NCSC launched a collaboration with the Thomas Reuters Institute to bolster the availability of AI tools, trainings, and recommendations for the courts.[155] The initiative is focused in particular on three areas: best practices in court and administration, AI for justice, and AI governance.[156] It offers an "AI Sandbox" in which judges and courts can experiment with different implementations of AI applications, alongside a database of judicial court orders, rules, and guidance documents from courts across the country.[157] This effort supports a broader community of practice, the "Court AI Implementers' Forum," through which state courts can collaborate on their AI implementations.[158] State bar associations and professional organizations across the country are also establishing committees to produce guidance on the use of AI; that which is directed at judges have emphasized a few points:[159] first, that AI use by the judiciary must be cabined—only judges can decide cases;[160] second, that judges have an ongoing ethical duty to understand and remain competent in technology;[161] and third, that while AI holds great promise, its risks must be managed.[162] For example, in October 2025, New York announced an AI policy for its courts, specifying which AI platforms may be used, and prohibited judges from "engag[ing] [AI] in the decision-

155. *AI in Courts Resource Center Launches to Empower Justice with AI*, THOMSON REUTERS (Jan. 6, 2025), https://www.thomsonreuters.com/en-us/posts/ai-in-courts/ai-in-courts-resource-center-launches/.

156. *Id.*

157. *AI in State Courts*, NAT'L CTR. FOR STATE CTS., https://www.ncsc.org/resources-courts/ai-state-courts (last visited June 9, 2025); *see also* Andre Assumpcao, *NCSC's 'Sandbox' Tool Aims to Help Courts Utilize AI Systems*, LEGALNEWS (Mar. 6, 2025), https://www.legalnews.com/Home/Articles?DataId=1582524.

158. *See Artificial Intelligence (AI)*, NAT'L CTR. FOR STATE CTS., https://www.ncsc.org/resources-courts/artificial-intelligence (demonstrating how platforms are encouraging court professionals to share their experience with AI).

159. *See* Marchant, *supra* note 81, at 483–85.

160. *Id.* (referring to the New Jersey Supreme Court's adoption of a Statement of Principles for the use of AI by the N.J. Courts); *see also* William M. Carlucci, Kaitlyn E. Stone & Michael Zogby, *Supreme Courts of Delaware and Georgia Take Steps to Regulate the Use of Artificial Intelligence*, NAT'L L. REV. (Oct. 24, 2024), https://natlawreview.com/article/supreme-courts-delaware-and-georgia-take-steps-regulate-use-artificial-intelligence (describing interim policy of the Delaware Supreme Court emphasizing that GenAI should not serve "as a substitute for judicial, legal, or professional expertise" or human "decision-making function[s]."); *see also Illinois Supreme Court Announces Policy on Artificial Intelligence*, ILL. CTS. (Dec. 18, 2024), https://www.illinoiscourts.gov/News/1485/Illinois-Supreme-Court-Announces-Policy-on-Artificial-Intelligence/news-detail/ (emphasizing that "[j]udges remain ultimately responsible for their decisions, irrespective of technological advancements").

161. *See* Marchant, *supra* note 81, at 484 (referencing the Michigan ethical advisory opinion).

162. *See id.* at 484–85 (referencing the Connecticut guidance, which includes an impact assessment methodology and the noting of risks by the New Jersey and Michigan statements).

making tasks a judge is ethically obligated to perform."[163] Meanwhile West Virginia's judicial commission concluded that while AI may be used for research purposes, it may not be used to reach a conclusion on the outcome of a case.[164]

In 2025, the Administrative Office of the Courts (AOC) launched a new "AI Pilot" initiative to work with district and circuit courts to identify and meet their needs using AI.[165] While the details are still emerging, chatbots that can enable the provision of services after hours appear to be one area of particular interest.[166] The AOC has also reportedly created a task force to ascertain the need for policies on the judicial use of AI.[167]

Notably in October 2025, following the news that two district court judges used generative AI to draft factually inaccurate court orders, Senate Judiciary Committee Chairman Chuck Grassley called for the judicial branch to develop more decisive, meaningful and permanent AI policies and guidelines.[168]

These broader, systemic efforts complement what one judge has described as the individualized journey of each chamber to explore how best to integrate AI into its workflows,[169] particularly with an increasing influx of clerks trained on and familiar with AI tools. This process requires a careful weighing of the benefits of AI against the panoply of risks AI technologies still present, with a number of judges deciding it is worth it to use AI for core tasks.[170] Below, we

163. Nikola L. Datzov, *AI Jurisprudence: Toward Automated Justice*, 23 NW. J. TECH. & INTELL. PROP. 1, 83 (2025).

164. *Id.*

165. *See* Madison Alder, *U.S. Court System Eyeing AI Use Cases for Access to Justice, Cost Savings*, FEDSCOOP (May 6, 2025), https://fedscoop.com/u-s-court-system-eyeing-ai-use-cases-for-access-to-justice-cost-savings/.

166. *Id.*

167. Jacqueline Thomsen, *US Courts Cautiously Experiment with AI to Speed Up Their Work*, BLOOMBERG L. (Apr. 7, 2025), https://news.bloomberglaw.com/us-law-week/us-courts-cautiously-experiment-with-ai-to-speed-up-their-work.

168. *Grassley Releases Judges' Responses Owning Up to AI Use, Calls for Continued Oversight and Regulation*, U.S. SENATE COMM. ON THE JUDICIARY (Oct. 23, 2025), https://www.judiciary.senate.gov/press/rep/releases/grassley-releases-judges-responses-owning-up-to-ai-use-calls-for-continued-oversight-and-regulation.

169. J. Herbert B. Dixon Jr., *I Am a Judge. Should I Use AI to Do My Job? Which AI Tools Should I Use?*, 64 JUDGES' J. 36 (2025), https://www.americanbar.org/content/dam/aba/publications/judges_journal/vol64no1-jj2025-tech.pdf.

170. *See id.*; *see also* James O'Donnell, *Meet the Early-Adopter Judges Using AI*, MIT TECH. REV. (Aug. 11, 2025), https://www-technologyreview-com.cdn.ampproject.org/c/s/www.technologyreview.com/2025/08/11/1121460/meet-the-early-adopter-judges-using-ai/amp/ (detailing the use of generative AI by individual US judges to summarize cases, generate

highlight selected instances in which U.S. courts have adopted AI, mindful that our description includes only a small fraction of the initiatives unfolding across the American judiciary.

1. Core Judicial Functions

One of the most intriguing ways in which AI has been used, in a few cases, is as a tool for engaging in legal reasoning. In a pair of decisions, Judge Kevin Newson of the 11th Circuit Court of Appeals described a considered experiment to use LLMs, not as drafting tools, but as thought partners of sorts.[171] At issue in a first case was the meaning of the word "landscaping," and whether or not a trampoline, installed at the ground-level, qualified.[172] Working with a clerk, Judge Newsom's concurrence describes asking ChatGPT and other LLMs about the ordinary meaning of the term, and then ultimately, for an analysis of the legal question.[173] On this basis, as well as a consideration of the benefits and drawbacks of using AI, the judge concluded that LLMs could be helpful in the judicial task of ordinary-meaning making. Alongside dictionaries, semantic canons, and other approaches, LLMs deserved their place, he concluded, in the "textualist toolkit."[174] In a second case, Judge Newsom extended the experiment to the task of interpreting the phrase, "physically constrained."[175] He found the LLMs useful for understanding composite phrases, which tend not to be listed in dictionaries, even if their responses varied each time the queries were asked.[176] In another case, before the D.C. District Court of Appeals, the majority, concurrence, and dissent each discussed the merits of using ChatGPT as a source for determining whether or not leaving a dog in a hot car amounted to animal cruelty, among other topics.[177] But in contrast to Judge Newsom, the judges expressed much greater

timelines, come up with questions for attorneys, and order information from complex dockets).

171. Snell v. United Specialty Ins. Co., 102 F.4th 1208, 1221 (11th Cir. 2024) (Newsom, J., concurring); United States v. Deleon, 116 F.4th 1260, 1270 (11th Cir. 2024) (Newsom, J., concurring).

172. *Snell*, 102 F.4th at 1212–13.

173. *Id.* at 1224–25.

174. *Id.* at 1226 (describing LLMs as "one implement among several in the textualist toolbox—to inform ordinary–meaning analyses of legal instruments."). Notably, the majority opinion took the opinion that it did not need to decide the meaning of the word "landscaping" in order to resolve the appeal. *Id.* at 1221.

175. *Deleon*, 116 F.4th at 1274 (describing references to three LLMs, each of which he posed his queries to ten times).

176. *Id.*

177. *See, e.g.*, Ross v. United States, No. 23-CM-1067, at *11 n.2 (D.C. Ct. App., Feb. 20, 2025), https://www.dccourts.gov/sites/default/files/2025-02/Ross-v-United-States-23-CM-1067-S.pdf; *see also id.* at *20–27 (Howard, J., concurring); *id.* at *37 n.5 (Deahl, J., dissenting).

skepticism that ChatGPT is necessarily "a good proxy for what is, and what isn't, common knowledge."[178]

In contrast to generative AI technology, risk assessment tools have been widely used by U.S. courts, in criminal justice cases. Such instruments use a variety of factors to estimate the probability of an outcome like reoffending or failing to appear in court. The Mapping Pretrial in Justice project has documented the use of pretrial risk assessment tools in all but four states,[179] the two most popular instruments being the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) and LSI-R (Level of Service Inventory-Revised).[180] As described above, the use of these instruments has been highly scrutinized and criticized; a recent study that probed public perception of the use of AI in bail and sentencing contexts found that people tended to view judges that relied solely on their own expertise more positively than judges who relied on AI, either partially or completely.[181] The mixed reception of the public to the use of such inputs in criminal justice contexts has likely contributed to the cautious approach taken by the courts to AI in general.

Courts are also experimenting with online dispute resolution (ODR)—the use of information and communications technology to help individuals resolve their disputes without having to resort to formal court processes. Algorithms are already reportedly being used as part of ODR proceedings in Utah, Wisconsin, California, and elsewhere, in a number of lower-stakes contexts.[182] In the future, one could imagine AI playing a number of roles in more complex proceedings, for example as a digital advisor helping guide parties through distinct phases of a dispute resolution process.[183]

178. *Id.* at *11 n.2.

179. *National Landscape*, MAPPING PRETRIAL INJUSTICE, https://pretrialrisk.com/national-landscape/.

180. *See* Simshaw, *supra* note 6, at 806 n.63.

181. *See* Fine et al., *supra* note 10, at 492. *But see id.* at 491 (finding Black study participants to have more positive views of judicial processes, whether or not with AI).

182. Samuel D. Hodge Jr., *Is the Use of Artificial Intelligence in Alternative Dispute Resolution a Viable Option or Wishful Thinking?*, 24 PEPP. DISP. RESOL. L.J. 91, 108 (2024) (describing the application of "algorithm-based ODR" to "small monetary disputes, traffic tickets, outstanding warrant issues, and 'low-conflict family court cases.'").

183. Kendal Enz & Colin Rule, *Nerding Out on Dispute Resolution: An Interview with ODR.com's Colin Rule*, MEDIATE (Jan. 8, 2025), https://mediate.com/nerding-out-on-dispute-resolution-an-interview-with-odr-coms-colin-rule/; Kate Shonk, *AI Mediation: Using AI to Help Mediate Disputes*, PROGRAM ON NEGOT.: HARV. L. SCH. (June 10, 2025), https://www.pon.harvard.edu/daily/mediation/ai-mediation-using-ai-to-help-mediate-disputes/ (describing a mediation in which ChatGPT was used by a mediator to suggest a number to propose to the parties, which was used to help break the impasse between them).

2. Court Management

Outside of the realm of core judicial functions, AI has been applied to a variety of court management tasks. Since 2016, the 15th Judicial Circuit (Palm Beach County) has used a combination of "narrow AI" and "robotic process automation (RPA)" to streamline docketing tasks.[184] A technology tool used by the county, Intellidact AI,[185] uses machine learning to automatically 'read' filed documents, extract relevant data, fill out docket sheets to be put into the case management system, and then publish the resulting documents.[186] Implementation of the system in Palm Beach was reported to lead to a 4-fold increase in processing speed for electronic filings, along with a significant reduction in errors and an increase in user satisfaction, and an approximate $2.5 million in savings annually.[187] Similar systems have been deployed in Texas and California.[188] Other experiments in court management include the use of AI to aggregate information from multiple sources in order to better manage juvenile court cases in Montgomery County, Ohio.[189]

Detecting legal errors on forms and pleadings represents another time-consuming and laborious task ripe for automation. Los Angeles Superior Court is working with researchers from Stanford on a tool that can check "default judgements," which are entered when a defendant fails to show up or respond to a complaint, and ensure that they are actually legally warranted.[190] Early results suggest that AI may be able to detect errors in up to 10% of cases as compared to a 1% error detection rate in the case of human review, a

184. *See* Glen Bischoff, *Key Takeaways from the 2023 Courts Technology Conference—Part 1*, MISSION CRITICAL PARTNERS (Oct. 5, 2023), https://resources.missioncriticalpartners.com/insights/key-takeaways-from-the-2023-courts-technology-conference-part-1 (discussing Parik Chokshi—circuit court clerk, comptroller, and director of enterprise applications for Palm Beach County, Florida).

185. *Palm Beach Clerk Receives National Digital Innovation Award for Its Use of CSI Intellidact AI*, FL. CT. CLERKS & COMPTROLLERS (Jan. 4, 2019), https://www.flclerks.com/news/news.asp?id=432529.

186. Felicity Bell, Lyria Bennett Moses, Michael Legg, Jake Silove & Monika Zalnieriute, *AI Decision-Making and the Courts: A Guide for Judges, Tribunal Members and Court Administrators*, AUSTRALASIAN INST. JUD. ADMIN. LTD. 26 (2022), https://aija.org.au/publications/ai-decision-making-and-the-courts-a-guide-for-judges-tribunal-members-and-court-administrators/.

187. *See* Bischoff, *supra* note 184.

188. *See* Bell et al., *supra* note 186.

189. *See* Marcus W. Reinkensmeyer & Raymond L. Billotte, *Artificial Intelligence (AI): Early Court Project Implementations and Emerging Issues*, NACM: CT. MANAGER (Aug. 2019), https://thecourtmanager.org/articles/artificial-intelligence-ai-early-court-project-implementations-and-emerging-issues/.

190. Shana Lynch, *Harnessing AI to Improve Access to Justice in Civil Courts*, HAI: STAN. U. HUM.-CENTERED A.I. (Mar. 4, 2025), https://hai.stanford.edu/news/harnessing-ai-to-improve-access-to-justice-in-civil-courts.

dramatic increase.[191] This can be particularly meaningful in the context of evictions where a landlord's improper service, damages miscalculations, or the failure to meet a rental arrears threshold could all be reasons that a default judgment is improper.

3. Interfacing with the Public

A number of U.S. courts have deployed bots—both physical and virtual—to provide triage and navigational services to the public. In this Section, we highlight a handful of them to provide a sense of the range of uses of bots by courts, as well as legal aid service providers.[192]

In an early effort, the 20th Circuit Court, Ottawa County, Michigan, introduced a robotic "concierge" at the local courthouse. Court Operated Robot Assistant (CORA) provided maps and directions, court dockets, judge biographies, and answers to frequently asked questions (FAQs), in Spanish and English,[193] eliciting a positive reception among certain court visitors, but, also, cost and job displacement concerns.[194]

Virtual chatbots have been more widely deployed, for example, by the Arizona Judicial branch, California Superior counties of Riverside and Los Angeles, as well as by courts in Montana and elsewhere.[195] In 2022, the 11th Judicial Circuit of Florida announced the launch of an online AI chatbot named SANDI (Self-Help Assistant Navigator for Digital Interactions).[196] Alongside website navigation assistance, SANDI offers assistance to people representing themselves in divorce and other Family Division cases.[197] Behind the visage of a digital avatar, and capable of receiving both voice and text commands, SANDI can answer frequently asked questions about the Family Court Self-Help Program and direct users to the appropriate web pages for forms and instructions.[198] The court reportedly experienced a 94% reduction

191. Stanford HAI, *HAI Seminar with David Engstrom: AI and Access to Justice*, at 42:55, YOUTUBE (Feb. 28, 2025), https://youtu.be/qS9CEdymWxI?si=Zf4rM0vZYCapGG-N.

192. The National Center for State Courts' 2024 report on Chatbots has compiled a list of them. *See* A. Souza & Z. Zarnow, *Court Chatbots: How to Build a Great Chatbot for Your Court's Website.*, NAT'L CTR. FOR STATE CTS. 25–26 (2024), https://www.ncsc.org/sites/default/files/media/document/Court-Chatbots.pdf.

193. Reinkensmeyer et al., *supra* note 189.

194. *Id.*

195. Souza et al., *supra* note 192.

196. *Miami-Dade Courts Now Offer Website Navigation Help Via Online Chat with Digital Assistant SANDI*, ELEVENTH JUD. CIR. FLA. (July 25, 2022), https://www.jud11.flcourts.org/Court-Announcements/ArtMID/584/ArticleID/4522/Miami-Dade-Courts-Now-Offer-Website-Navigation-Help-via-Online-Chat-with-Digital-Assistant-SANDI.

197. *Id.*

198. Eunice Sigler, *SANDI: Improving Court Access and Service in Miami with an Advanced Artifical Intelligence Chatbot*, CT. NEWS FLA. (June 28, 2023), https://news.flcourts.gov/All-

in live chats after adopting the technology.[199] The bot's expertise also "grew" based on interactions with the public, synthesizing answers to new questions built upon an initial knowledge base.[200] Using a similar technology, New Mexico implemented CLARA, a multilingual AI avatar stationed in courthouse kiosks[201] and online to assist the public. CLARA interacts through an on-screen persona and can answer questions or guide users to services in multiple languages, entered through text or voice command inputs.[202] These AI assistants can provide 24/7 services, in multiple languages, in a form much more friendly than court websites.

As of the time of this writing, the Alaska Court System (ACS) was working to develop an AI-powered chatbot called the Alaska Virtual Assistant, or AVA, with the legal tech firm LawDroid.[203] The goal is to create a conversational interface for delivering information currently captured within the 220 pages of static content on the Court's website.[204] A presentation of the project reported on the need for deliberation, vetting, and testing while the model was fine-tuned.[205]

In 2025, the Supreme Court of Nevada's Administrative Office of the Courts enlisted the help of technology company CiviLaw.Tech to develop online tools including instructional step-by-step guides, informative videos about navigating the court system, and an AI-powered chatbot that offers guidance in over 50 languages.[206]

Court-News/SANDI-Improving-Court-Access-and-Service-in-Miami-with-an-Advanced-Artificial-Intelligence-Chatbot.

199. *Id.*

200. *Id.*

201. *NM Ranked #1 in Nation for Language Access in the Justice System*, N.M. ADMIN. OFF. CTS. (June 15, 2021), https://nmcourts.gov/wp-content/uploads/2024/03/NM-ranked-1-in-nation-for-language-access-in-the-justice-system-June-15-2021.pdf.

202. *See* Souza et al., *supra* note 192, at 18 (describing Clara as "speak-to-chat").

203. Natalie Runyon, *Chatbots for Justice: Building AI-Powered Legal Solutions Step By Step*, THOMSON REUTERS (Mar. 12, 2025), https://www.thomsonreuters.com/en-us/posts/ai-in-courts/chatbots-for-justice-building-ai-powered-legal-solutions/.

204. Jeannie Sato, Dir., Access to Just. Servs., & Tom Martin CEO, LawDroid, Tech for All: Applications of AI to Increase Access to Justice 2, NAT'L CTR. FOR STATE CTS., https://nationalcenterforstatecourts.app.box.com/s/aghv4c5h169wdysq74n04lt5nbgxep0s. Hawai'i has introduced a similar AI tool for navigating court website information. *See Hawai'i State Judiciary Launches AI-Powered KolokoloChat for Law Day 2025*, HAW. STATE JUDICIARY (May 1, 2025), https://www.courts.state.hi.us/news_and_reports/2025/05/hawai%CA%BBi-state-judiciary-launches-ai-powered-kolokolochat-for-law-day-2025.

205. *See generally* Sato & Martin, *supra* note 204.

206. *Nevada Judiciary Expands Free Legal Resources with Self-Help Website*, NEV. ADMIN. OFF. OF CTS. (Mar. 3, 2025),
https://nvcourts.gov/aoc/aoc_news/
nevada_judiciary_expands_free_legal_resources_with_self-help_website.

Adjacent to the formal legal system, legal aid organizations have also developed a number of chatbots to help unrepresented litigants exercise their rights. For example, in 2024, Legal Aid of North Carolina (LANC) developed a generative AI chatbot (named "LIA") that provides answers to legal questions in English and Spanish.[207] Developed in LANC's Innovation Lab in collaboration with LawDroid, LIA is designed to efficiently provide high-quality legal information to underserved communities. LIA automates routine communications and provides self-service options for simple legal matters, streamlining the overall client experience. Powered by models like GPT-4 and BERT, and supported by LawDroid's technical infrastructure, LIA focuses on high-demand areas such as domestic violence, child custody, landlord-tenant issues, and consumer law. Meanwhile, in Missouri, an online screening tool helps tenants determine eligibility for legal assistance before connecting with program staff. The AI-chatbot, "MOLS," can help individuals determine whether their issue is one legal aid can address.[208] Also for tenants, Rentervention is an AI virtual assistant launched by the Law Center for Better Housing, the Illinois Equal Justice Foundation, and the Lawyers Trust Fund of Illinois.[209] Through it, Illinois renters can access information and resources on housing rights, as well as connect with an attorney if legal advice is needed. A counterpart tool in New York, "Roxanne," developed in partnership with New York University School of Law and the legal automation company Josef assists tenants in addressing housing repair issues. The tool seeks to both educate renters and help them enforce their rights.[210]

These collaboratively developed tools demonstrate the responsiveness of the courts and legal community to the needs of the public and underserved litigants. But the extent to which the initiatives described above remain "demonstration" projects as opposed to the norm in U.S. courts depend on a number of factors beyond the scope of this Article including the emergence

207. *Legal Aid of North Carolina Launches LIA 2.0, Marking a New Era in Accessible, AI-Powered Legal Information*, LEGAL AID OF N.C. (Nov. 18, 2025), https://legalaidnc.org/2025/11/18/legal-aid-of-north-carolina-launches-lia-2-0-marking-a-new-era-in-accessible-ai-powered-legal-information/.

208. *Places to Get Help*, MO. TENANT HELP, https://motenanthelp.org/places-to-get-help/.

209. Shiva Kooragayala, *Using Generative A.I. to Expand Legal Aid: The Case of "Rentervention"*, MEDIUM (July 17, 2024), https://medium.com/justice-rising/using-generative-a-i-to-expand-legal-aid-the-case-of-rentervention-88df92e477c1.

210. Bob Ambrogi, *AI-Powered Tool Launches to Help New York Tenants Enforce Their Repair Rights*, LAWSITES (Jan. 7, 2025), https://www.lawnext.com/2025/01/ai-powered-tool-launches-to-help-new-york-tenants-enforce-their-repair-rights.html; *see also* Colleen V. Chien & Miriam Kim, *Generative AI and Legal Aid: Results from a Field Study and 100 Use Cases to Bridge the Access to Justice Gap*, 57 LOY. L.A. L. REV. 903, 967–68 (2025) (describing, among others, Rasa and Visalaw.Ai).

of funding models and development of scalable solutions within a fragmented landscape. An additional, formative factor in the case of legal aid technologies will be the enduring strength of laws prohibiting the unauthorized practice of law, which limit the support that AI tools can offer to the provision of legal information, not advice.[211]

IV. THE PATH FORWARD: INTEGRATING ARTIFICIAL INTELLIGENCE IN JUDICIAL SYSTEMS

Our brief survey of the use of artificial intelligence (AI) by three judicial systems underscores that successful implementation of AI transcends mere technological adoption. Rather, meaningful AI deployment requires consideration of judicial operations, data governance, and justice delivery mechanisms and social imperatives—harmonizing technological capabilities with institutional capacity and jurisprudential traditions.

A. DATA INFRASTRUCTURE AND INSTITUTIONAL FOUNDATIONS

The experiences of Brazil and China demonstrate how robust data architecture can serve as the cornerstone for scalable AI utilization within judicial frameworks. Brazil's commitment to digital recordkeeping, culminating in near-universal electronic filing protocols and the establishment of the CODEX data repository show how digitization constitutes merely the preliminary phase of technological integration. The substantive challenge lies in developing interoperable, high-fidelity datasets capable of supporting sophisticated analytical applications and predictive modeling systems. In a similar vein, China's comprehensive national Big Data initiatives have enabled platforms such as the Same Type Court Reference (STCR) system and the FaXin judicial database, which integrate hundreds of millions of legal instruments into unified platforms. Atop these data infrastructures, meaningful JudicialTech has been developed in each country to effectively provide a "first draft" of a wide variety of court documents. Both jurisdictions successfully transitioned from isolated pilot programs to systemic AI integrations because they coupled technical preparedness with sustained institutional commitment and strategic vision.

In contrast, the United States experience exemplifies how a more cautious approach to automation, centered in due process and individualized justice as well as decentralized experimentation can also support innovation in service of judicial mission, individual court preference, and autonomy, but at a much smaller scale. As the efforts described earlier to centralize and coordinate

211. *See* Stephanos Bibas, *Lawyers' Monopoly and the Promises of AI*, 134 YALE L.J. F. 920, 921 (2025).

across the U.S. judiciary take shape, the likelihood of greater technological and procedural legal interoperability—essential for more systematic reform—will also increase.

B. GOVERNANCE FRAMEWORKS AND JUDICIAL LEADERSHIP

Our comparative analysis also reveals that diverse governance models may successfully foster technological advancement, once there exists committed leadership dedicated to responsible innovation principles. Brazil's centralized oversight mechanism, administered through the National Council of Justice, facilitated the development of rights-based regulatory frameworks, including Resolutions No. 332/2020 and 615/2025, which effectively balance technological innovation with constitutional safeguards and due process protections. China's hierarchical, state-directed implementation model has allowed it to develop initiatives with local and national reach. The United States experience, while characterized by jurisdictional fragmentation, also demonstrates how localized leadership and professional organizations can advance AI adoption even in the absence of formal federal mandates or comprehensive regulatory frameworks.

These divergent approaches confirm that no singular model guarantees successful implementation. The determinative factors extend beyond technological sophistication to encompass institutional vision and organizational capacity to integrate AI systems in alignment with existing legal frameworks and public expectations regarding judicial administration.

C. BALANCING TECHNOLOGICAL INNOVATION, HUMAN ADJUDICATION, AND SOCIETAL IMPERATIVES

Despite significant differences in their approaches, Brazil, China, and the United States commonly confront the fundamental challenge of balancing technological innovation with the preservation of human judicial discretion and the protection of constitutional and statutory rights. Brazil's constitutional framework explicitly mandates that AI systems function as auxiliary tools rather than substitutes for judicial decision-making processes. China's automated systems, while more extensively integrated, continue to require human oversight in critical adjudicatory functions, though the equilibrium between algorithmic processing and human intervention differs substantially from Western models. In the United States, early experiences with risk assessment tools as well as more recent cases of litigant uses of generative AI have collectively illustrated the importance of human oversight. In the absence of centralized mandates, they have increased the pressure on local court governance, professional responsibility codes, and judicial discretion. The societal pressures driving AI adoption—including massive litigation volumes, acute shortages of legal professionals, and persistent access-to-justice gaps—

underscore the urgency of achieving this delicate balance. However, sustainable integration requires more than technical solutions: it demands institutional wisdom to deploy AI responsibly, ensuring that technological systems enhance rather than supplant the core adjudicatory functions of judicial institutions.

V. CONCLUSION

The comparative experiences of Brazil, China, and the United States illustrate that AI can meaningfully reshape judicial administration through diverse institutional pathways. Each jurisdiction reflects distinct governance models, legal traditions, and societal priorities, which in turn inform their respective approaches to AI integration. Rather than suggesting a singular trajectory or ranking of advancement, these variations highlight the contextual nature of technological adaptation within judicial systems.

Common principles nonetheless emerge: the importance of robust data infrastructure, thoughtful and accountable leadership, human-centered design, and the need to align technological capabilities with institutional mandates and values. These shared elements underscore that while implementation strategies may differ, foundational challenges—and aspirations—remain broadly convergent.

As courts move from pilot initiatives toward more enduring forms of AI integration, a central insight is that what matters is not only the relative speed or scale of adoption, but the quality and care with which AI is embedded into judicial processes. As with many things, automated justice can be considered a "double-edged sword"—easing the load of judges on the one hand but reducing incentives for the thorough consideration of the record and identification of opportunities to evolve the law.[212] A higher volume of cases handled with AI does not necessarily equate to a greater number of people receiving adequate help or the increased legitimacy of the courts.[213] Technology offers powerful tools for institutional improvement, but its contribution depends on transparent, deliberate governance that centers humans—in the tasks of both formulating justice as well as receiving it. The future of AI in the judiciary will be determined not by the sophistication of algorithms, but by the evolutionary capacity of legal institutions to integrate them in ways that strengthen rather than substitute the foundational promise of justice under law.

212. *See* Li, *supra* note 116, at 5–7.
213. *Id.* at 7.

THE OMB ARTIFICIAL INTELLIGENCE MEMORANDA

Sorelle A. Friedler[†] *& Andrew D. Selbst*[††]

ABSTRACT

Under the Biden and Trump Administrations, the Office of Management and Budget issued two memoranda on the use of artificial intelligence (AI) by the federal government. The memos set out minimum required risk management practices and associated governance structures that must be in place within federal government agencies before AI can be used. This Article traces the history of the OMB AI memos, explaining their shared origin in a decade of advocacy within civil society, industry, and academia that led to the creation of the Blueprint for an AI Bill of Rights by the Biden Administration's Office of Science and Technology Policy, which then fed directly into the Biden AI Memo, before it was replaced by the Trump Administration's version.

The Article then makes two arguments about the significance these memos. First, the lineage of the memos reveals the concern with practical implementation of minimum practices and safeguards in order to protect civil rights. Perhaps surprisingly, while the Trump Administration's replacement reflects the updated priorities of the new Administration, it keeps much of the structure and substance of the original memo, including some of the civil rights orientation and the requirement that an agency must meet the minimum practices or cease using the AI. Second, these memos serve an important but rarely recognized regulatory role within the government as what we call "intermediate instruments." By describing requirements at a level of specificity that makes them actionable while at a level of generality to make them applicable across many agencies and use cases, these memos become necessary governance tools that bridge the principles expressed in executive orders and the day-to-day practice of agencies. Such intermediate instruments are not often recognized as important in

---

DOI: https://doi.org/10.15779/Z38Z892H8Q

† Shibulal Family Professor of Computer Science, Haverford College. Friedler served as the Assistant Director for Data and Democracy in the Biden White House Office of Science and Technology Policy, where she was one of the authors of the AI Bill of Rights and contributed to the Biden AI Memo discussed in this Article.

†† Professor of Law, University of California, Los Angeles, School of Law. The authors would like to thank the organizers and participants of the 2025 BTLJ-BCLT Spring Symposium: AI Governance at the Crossroads, especially including Olivia Zhu, for helpful feedback.

themselves, but the Article argues that they are worthy of independent recognition because they are likely widely used in oversight schemes of distributed bureaucratic structures.

TABLE OF CONTENTS

I. INTRODUCTION

On March 28, 2024, the Biden Administration's Office of Management and Budget (OMB) issued a memorandum on artificial intelligence (AI) ("Biden AI Memo").[1] On April 3, 2025, OMB, now under President Trump, replaced it with one that reflected the new Administration's priorities ("Trump AI Memo").[2] In this Article, we aim to explain the history and significance of these two documents. We trace a historical though-line beginning from civil society, academic, and government work on civil rights harms stemming from AI to the concrete protections enacted by the OMB Memos. In addition to their significance as civil rights protections, we argue that these memos are a

1. OFF. OF MGMT. & BUDGET, EXEC. OFF. OF THE PRESIDENT, M-24-10, ADVANCING GOVERNANCE, INNOVATION, AND RISK MANAGEMENT FOR AGENCY USE OF ARTIFICIAL INTELLIGENCE (Mar. 28, 2024), https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf [https://perma.cc/F8JB-X7S2] [hereinafter Biden AI Memo].

2. OFF. OF MGMT. & BUDGET, EXEC. OFF. OF THE PRESIDENT, M-25-21, ACCELERATING FEDERAL USE OF AI THROUGH INNOVATION, GOVERNANCE, AND PUBLIC TRUST (Apr. 3, 2025), https://www.whitehouse.gov/wp-content/uploads/2025/02/M-25-21-Accelerating-Federal-Use-of-AI-through-Innovation-Governance-and-Public-Trust.pdf [https://perma.cc/CCT8-2GYQ] [hereinafter Trump AI Memo].

model of what we call "intermediate instruments"—governance tools that are necessary to convert high-level governance principles into actionable structures, procedures, and requirements within federal agencies.

OMB is the executive branch office responsible for carrying out the President's policies for internal operations of the federal government. Though the actual organizational chart is complicated—and surprisingly difficult to locate[3]—in practice, OMB is divided into budget, management, and legislative review arms. The management side is a set of five statutory offices that report to the Deputy Director for Management (DDM).[4] The DDM is responsible for "[c]oordinating and supervis[ing] the general management functions" of the OMB including, but not limited to, "managerial systems, including the systematic measurement of performance," "procurement policy," and "information and statistical policy."[5] The DDM is also charged with "[p]roviding leadership in management innovation through . . . the adoption of modern management concepts and technologies."[6] To accomplish these objectives, the OMB regularly issues management memoranda to implement internal policies for technology adoption and procurement. The Biden AI Memo and Trump AI Memo (collectively the "OMB AI Memos") are two such memos.

The OMB AI Memos created binding guidance for federal agencies on how to address the risks of harm posed by government use of AI. The three main sections of the memos describe (1) a governance structure for AI within federal agencies, including the introduction of Chief AI Officers (CAIOs), (2) the development of agency AI strategies and other mechanisms to encourage innovation, and (3) required minimum risk management practices for federal government use of AI. In this Article, we focus on (3), the structure and incorporation of these minimum risk management practices, which set the

3. The current White House website does not list an organizational chart. *See Office of Management and Budget*, WHITE HOUSE, https://www.whitehouse.gov/omb/ [https://perma.cc/H3RH-ETZE] (last visited Nov. 11, 2025). The only organizational chart we can find is from the Obama White House. *See Organizational Chart*, OBAMA WHITE HOUSE ARCHIVES, https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/about_omb/omb_org_chart_0.pdf [https://perma.cc/P47Z-SCAD].

4. These offices are the Office of Information and Regulatory Affairs (OIRA), the Office of Federal Financial Management (OFFM), the Office of Federal Procurement Policy (OFPP), the Office of the Federal Chief Information Officer (OFCIO, serving the role of the E-Gov office), and the office of the Intellectual Property Enforcement Coordinator (IPEC). Each administration may create an organizational chart that additionally has other non-statutory offices report to the DDM, such as the U.S. Digital Service or the Made in America Office. *See* 31 U.S.C. § 503.

5. *Id.* § 503(b)(1), (b)(2)(A), (b)(2)(B), (b)(2)(D).

6. *Id.* § 503(b)(6).

memos up as key intermediate instruments for AI regulation. The minimum risk management practices are requirements that all agencies must follow in order to use AI; agencies not in compliance with the requirements are directed to cease use of those systems until they are brought into compliance. The memos also direct the CAIOs to oversee agency compliance, manage any necessary waivers or timeline extensions, and ensure reporting on agency use of AI to both OMB and the public by issuing AI Use Case Inventories.

These memos are most obviously important because they offer direct actionable guidance to federal agencies on civil rights protections for AI. But we argue here that they carry a separate conceptual importance; they are necessary intermediate instruments that convert principles into practice. In order to be effective, OMB guidance must be specific enough for agencies to follow and for OMB to certify compliance, while also being general enough to apply across widely varied uses of AI across agencies, from detecting invasive bullfrogs[7] to veteran suicide risk assessment[8] and government service chatbots.[9] The OMB AI Memos do just that, by creating sufficiently specific requirements for AI governance, but at a level of generality that will enable agencies to adapt the guidance to their individual needs.

When the Trump Administration took over from the Biden Administration, it replaced the memos governing AI. It did so in part because the Biden Administration's focus on equity was not in line with the new Administration's goals. But given the political context, what is most striking about the replacement memo is not the differences, but rather the continuity. The government needs documents like this simply to function, and while the policies did change between administrations, the governance structures and strategies largely did not. This demonstrates that the OMB AI Memos are an important anchoring step in the governance of AI, not only internally to the government, but also for broader AI policy. They forge a path from the high-level principles that had previously dominated the AI governance landscape to practical implementation in varied contexts.

This Article proceeds in five parts. Part II of the Article describes the content of the OMB AI Memos—the definitions of AI, the AI systems that are covered, and the substantive requirements.

---

7. U.S. DEP'T OF THE INTERIOR, ARTIFICIAL INTELLIGENCE (AI) USE CASE INVENTORY, https://www.doi.gov/ai/use-case-inventory [https://perma.cc/W8W5-KKPB].

8. U.S. DEP'T OF VETERANS AFFS., AI USE CASE INVENTORY, https://department.va.gov/ai/ai-use-case-inventory/ [https://perma.cc/29PX-3MZP].

9. U.S. DEP'T OF HOMELAND SEC., ARTIFICIAL INTELLIGENCE USE CASE INVENTORY LIBRARY, https://www.dhs.gov/publication/ai-use-case-inventory-library [https://perma.cc/7777-XJSY].

Part III tells the story of the Biden AI Memo, beginning with its origins in a decade of work in civil society, academia, and government, which ultimately converged on high-level AI principles and led to the Blueprint for an AI Bill of Rights ("AI Bill of Rights"), a white paper produced by the White House Office of Science and Technology under the Biden Administration. Part III then explains the influence of the AI Bill of Rights on the resulting Biden AI Memo, drawing on language in both to demonstrate the direct relationship.

Part IV discusses the Trump AI Memo. While it is unsurprising that the new Administration replaced and updated the memo to reflect new priorities, this Part argues that what is perhaps more significant is the degree of continuity between the memos. We argue that the two biggest changes in the Trump AI Memo are (1) a more explicit move to a risk regulation framework and (2) the deprioritization of bias and equity issues. But the overall governance approach remains the same. The requirement of minimum risk management practices and the deference to agency expertise live on in the new memo, despite policy positions that were reversed or changed.

Part V argues that the importance of both OMB AI Memos collectively is to support the transition from principles to practice. Whenever a large organization, such as the federal government, endeavors to govern AI based on principles across many different units and domains, it must have some form of intermediate instrument to make this implementation possible. Despite the differences between administrations, the lesson of the OMB AI Memos is that the creation of implementation instructions has importance beyond the specific policies implemented.

Part VI concludes by discussing the lasting legacy of the Biden AI Memo.

II. THE OMB AI MEMOS

OMB was directed to produce the memos by three sources: the AI in Government Act of 2020,[10] the Advancing American AI Act,[11] and President Biden's (now rescinded) Executive Order (EO) 14110.[12] The AI in Government Act required that OMB create guidance for federal agencies through a process including public input to mitigate against bias or "any unintended consequences" resulting from government use of AI systems.[13]

10. AI in Government Act, Pub. L. No. 116–260, 134 Stat. 2286 (2020) (codified at 40 U.S.C. § 11301 note).

11. Advancing American AI Act, Pub. L. No. 117-263, 136 Stat. 3668 (2022) (codified at 40 U.S.C. § 11301 note).

12. Exec. Order No. 14110 on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, 88 Fed. Reg. 75191 (2023) [hereinafter EO 14110].

13. AI in Government Act, *supra* note 10, at 2288, § 104(a)(3).

The Advancing American AI Act essentially reiterated the order, as there was not yet a memo by 2022, adding that the memo should draw on the principles in EO 13960,[14] President Trump's executive order on AI signed in the waning days of the first Trump Administration. Finally, President Biden's EO 14110, signed in late 2023, required that OMB develop minimum risk management practices and specifically scoped this requirement to focus on "government uses of AI that impact people's rights or safety."[15]

The resulting OMB AI Memos lay out governance practices and requirements for federal government use of AI,[16] including both internal government development of AI systems and agencies' use of privately developed systems via procurement.[17]

A. DEFINITIONS AND COVERAGE

The OMB AI Memos require a set of specific agency practices based on the type and use case of "covered AI."[18] They define AI according to the John S. McCain National Defense Authorization Act for Fiscal Year 2019[19] to include the following:

---

14. Advancing American AI Act, *supra* note 11, at 3670, § 7224(a)(2).

15. EO 14110, *supra* note 12.

16. Trump AI Memo, *supra* note 2, *passim*. The equivalent language in the Biden AI Memo stated that it created "agency requirements and guidance for AI governance, innovation, and risk management." Biden AI Memo, *supra* note 1, at 1. The memos also contain directives to enhance innovation in government use of AI. *Id.* We are only interested in the oversight and risk management role of these memos, so we set aside the part of the memos designed to promote AI use in government.

17. Biden AI Memo, *supra* note 1, at 3 ("This memorandum provides requirements and recommendations that, as described in more detail below, apply to new and existing AI that is developed, used, or procured by or on behalf of covered agencies."); Trump AI Memo, *supra* note 2, at 4 ("This memorandum provides requirements and recommendations that apply to new and existing AI that is developed, used, or acquired by or on behalf of covered agencies."). Separate memoranda describe the specific contract requirements that the government must follow to procure AI systems, but those memos in turn point back to the OMB AI Memos for the substantive minimum required practices. OFF. OF MGMT. & BUDGET, EXEC. OFF. OF THE PRESIDENT, M-24-18, ADVANCING THE RESPONSIBLE ACQUISITION OF ARTIFICIAL INTELLIGENCE IN GOVERNMENT (Sep. 24, 2024), at 4–5, 8 ("Agencies must ensure their AI acquisitions comply with the risk management requirements identified in OMB Memorandum M-24-10 if the AI is used in a way that impacts rights or safety, while also continuing to prioritize privacy, security, data ownership, and interoperability"); OFF. OF MGMT. & BUDGET, EXEC. OFF. OF THE PRESIDENT, M-25-22, DRIVING EFFICIENT ACQUISITION OF ARTIFICIAL INTELLIGENCE IN GOVERNMENT (Apr. 3, 2025), at 10 ("Contracts must ensure compliance with minimum risk management practices for high-impact use cases as required under M-25-21.").

18. Trump AI Memo, *supra* note 2, at 4; Biden AI Memo, *supra* note 1, at 3.

19. Biden AI Memo, *supra* note 1, at 26–27 (quoting Pub. L. No. 115-232, § 238(g)); Trump AI Memo, *supra* note 2, at 18. This definition was, in turn, the scoping definition used

> 1. Any artificial system that performs tasks under varying and unpredictable circumstances without significant human oversight, or that can learn from experience and improve performance when exposed to data sets.
>
> 2. An artificial system developed in computer software, physical hardware, or other context that solves tasks requiring human-like perception, cognition, planning, learning, communication, or physical action.
>
> 3. An artificial system designed to think or act like a human, including cognitive architectures and neural networks.
>
> 4. An artificial system designed to think or act like a human, including cognitive architectures and neural networks.
>
> 5. An artificial system designed to act rationally, including an intelligent software agent or embodied robot that achieves goals using perception, planning, reasoning, learning, communicating, decision making, and acting.

The memos add "technical context" to guide the interpretation of what is covered by this definition of artificial intelligence. Two points are particularly notable: (1) the definition includes "machine learning (including deep learning as well as supervised, unsupervised, and semi-supervised approaches), reinforcement learning, transfer learning, and generative AI," and (2) "no system should be considered too simple to qualify as covered AI due to a lack of technical complexity (e.g., the smaller number of parameters in a model, the type of model, or the amount of data used for training purposes)."[20] These clarifications are important because they ensure that machine learning models like linear and logistic regression, sometimes used for high-impact systems, are considered within scope, along with more complex models frequently understood as the cause for concern.[21] "Covered AI," in turn, excludes use cases outside of an agency's core functions—for example, systems used solely for research or assessments of the AI itself in preparation for regulatory enforcement.[22]

---

by the legislation directing OMB to write the memos (both the AI in Government Act of 2020 and the Advancing American AI Act of 2022).

20. Trump AI Memo, *supra* note 2, at 18; Biden AI Memo, *supra* note 1, at 27.

21. Such simple, yet high-impact, systems include a VA health care risk assessment logistic regression model described in the AI use case inventory. *See* supra note 8, U.S. DEP'T OF VETERANS AFFS., AI Use Case Inventory (describing the VA's "Care Assessment Needs (CAN) Score" system: "CAN is a set of risk-stratifying logistic regression models run on Veterans receiving health care through VHA.").

22. Trump AI Memo, *supra* note 2, at 4; Biden AI Memo, *supra* note 1, at 3.

For covered AI, the OMB AI Memos offer additional distinctions that dictate what requirements apply. Under the Biden AI Memo, systems are classified as "safety-impacting" or "rights-impacting." Safety-impacting systems are defined to include covered AI with significant impacts on human life, climate and the environment, critical infrastructure, or strategic assets or resources. Rights-impacting systems are defined as those covered AI systems which have a significant effect on an individual's or entity's civil rights, civil liberties, privacy, equal opportunities, or access to resources or services.[23] Under the Trump AI Memo, these categorizations are collapsed into a single definition of "high-impact" AI, that includes covered AI with a significant effect on: civil rights, civil liberties, or privacy; individual's or entity's access to education, housing, critical government resources or services, or other programs; human health and safety; critical infrastructure or public safety; or strategic assets or resources.[24] Though the structure of the categories differs between the memoranda, the set of covered AI remains largely consistent.

In addition to providing definitions for high-impact, safety-impacting, and rights-impacting AI, both memoranda provide agencies with a specific list of AI use cases that are presumed included within these definitions.[25] Beyond the substantive force of the presumption, the list also serves an important communicative purpose to agencies, creating a government-wide understanding of which systems are high-impact. Specific systems presumed by both memoranda to be high-impact include "emergency services," "the medically relevant functions of medical devices," law enforcement risk assessments, access controls for benefits systems, and many others.[26] The Trump AI Memo removed some systems that had been on the Biden lists including "[m]aintaining the integrity of elections and voting infrastructure" and a wide variety of education-related AI systems.[27]

23. Biden AI Memo, *supra* note 1, at 29.

24. Trump AI Memo, *supra* note 2, at 19.

25. Trump AI Memo, *supra* note 2, at 21–22; Biden AI Memo, *supra* note 1, at 31–33.

26. Trump AI Memo, *supra* note 2, at 21–22; Biden AI Memo, *supra* note 1, at 31–33.

27. Education-related systems included in the Biden AI Memo as systems that are presumed to be rights-impacting, and excluded entirely from the Trump AI Memo's list, are: "In education contexts, detecting student cheating or plagiarism; influencing admissions processes; monitoring students online or in virtual-reality; projecting student progress or outcomes; recommending disciplinary interventions; determining access to educational resources or programs; determining eligibility for student aid or Federal education; or facilitating surveillance (whether online or in-person)." *Compare* Biden AI Memo, *supra* note 1, at 32, *with* Trump AI Memo, *supra* note 2, at 19.

B. SUBSTANTIVE PROVISIONS

The OMB AI Memos set out three high-level, substantive risk management components for agencies to accomplish: (1) building a governance structure for AI use, (2) reporting about AI uses via the AI Use Case Inventory, and most importantly, (3) implementing certain minimum risk mitigation requirements, without which the AI system cannot be used.

The governance structure centers on the creation of a Chief AI Officer (CAIO) position.[28] Bigger agencies—specifically those that are statutorily required to have a Chief Financial Officer—must additionally have an AI governance board.[29] CAIOs are responsible for oversight of the memoranda requirements, including any waivers granted from OMB's required practices[30] and for communication within the government about agency AI procedures and with the public via the AI Use Case Inventory.[31]

The AI Use Case Inventory was mandated by EO 13960[32] and the Advancing American AI Act,[33] and integrated into the Office of the Federal Chief Information Officer's Integrated Data Collection process under the Biden AI Memo.[34] It serves as a key component of the public's transparency into AI use by the federal government by requiring public visibility via a website into current and planned agency AI uses.[35] Under the instructions

28. Trump AI Memo, *supra* note 2, at 10–11; Biden AI Memo, *supra* note 1, at 4–8.

29. Trump AI Memo, *supra* note 2, at 11; Biden AI Memo, *supra* note 1, at 8–9.

30. Trump AI Memo, *supra* note 2, at 15; Biden AI Memo, *supra* note 1, at 17.

31. Trump AI Memo, *supra* note 2, at 10–11; Biden AI Memo, *supra* note 1, at 6.

32. Exec. Order No. 13960 on Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government, 85 Fed. Reg. 78939 (2020) [hereinafter EO 13960].

33. Advancing American AI Act, *supra* note 11, at 3670, § 7224(a)(2).

34. The Office of the Federal Chief Information Officer (OFCIO) is housed under the DDM and manages governmentwide IT policy including the Integrated Data Collection (IDC) process which is a quarterly data reporting mechanism from agency CIOs to OMB. *See Office of the Federal Chief Information Officer*, WHITE HOUSE, https://bidenwhitehouse.archives.gov/omb/management/ofcio/ [https://perma.cc/P5LB-ZP5A] (describing OFCIO); *Integrated Data Collection*, CIO.GOV, https://www.cio.gov/handbook/reporting/idc/ [https://perma.cc/FWG3-S585] (describing the IDC); and Biden AI Memo, *supra* note 1, at 5 (requiring that AI Use Case Inventories be collected via the IDC).

35. Biden AI Memo, *supra* note 1, at 5. The Biden Administration made the consolidated AI use case inventory available at https://web.archive.org/web/20250116141511/https://ai.gov/ai-use-cases/. Individual agency inventories are available at agency-maintained sites, see, e.g., https://www.dhs.gov/publication/ai-use-case-inventory-library [https://perma.cc/7777-XJSY]; https://www.hhs.gov/programs/topic-sites/ai/use-cases/index.html [https://perma.cc/FUE9-X8AY]; https://department.va.gov/ai/ai-use-case-inventory/ [https://perma.cc/29PX-3MZP].

stemming from the Biden AI Memo,[36] such reporting further included information about the assessed risks of the system and steps taken at the direction of the memorandum to mitigate those risks. It is not yet clear whether reporting from the Trump AI Memo will maintain these instructions, or revert to the structure of earlier inventories that did not include information about risks.[37]

Substantively, for those covered AI use cases determined to be high-impact and not waived from some or all requirements via a process with the CAIO, the memoranda lay out "minimum risk management practices." The requirements include both provisions that must be in place before use, such as pre-deployment testing and risk mitigation, and those that continue throughout the life of the system, such as ongoing performance monitoring. In addition to technical practices like performance assessments, the minimum risk management practices also include sociotechnical practices that take into account the humans operating and impacted by the system. These include "training, assessment, and oversight for operators of AI"[38] and incorporating feedback from the public. In some cases, these minimum practices are strict requirements, such as the requirement for pre-deployment testing,[39] while others, such as the practice of incorporating feedback from the public, are guidance to be applied by agencies "where appropriate."[40] The memos direct that they are to be implemented for any high-impact covered AI or agencies must cease using the system.[41]

The next Part discusses these requirements in greater depth after a discussion of their genesis.

---

36. OFF. OF THE FED. CIO, GUIDANCE FOR 2024 AGENCY ARTIFICIAL INTELLIGENCE REPORTING PER EO 14110 (Aug. 14, 2024), https://www.cio.gov/assets/resources/2024-Guidance-for-AI-Use-Case-Inventories.pdf [https://perma.cc/5G6A-4EPE].

37. *See, e.g.*, the results of the 2023 AI Use Case Inventory: https://github.com/ombegov/2024-Federal-AI-Use-Case-Inventory/blob/main/data/2023_consolidated_ai_inventory_raw.csv [https://perma.cc/BR9R-EV6E].

38. Trump AI Memo, *supra* note 2, at 17; Biden AI Memo, *supra* note 1, at 20 ("operators of *the* AI").

39. Trump AI Memo, *supra* note 2, at 15; Biden AI Memo, *supra* note 1, at 18.

40. Trump AI Memo, *supra* note 2, at 17; Biden AI Memo, *supra* note 1, at 22.

41. Trump AI Memo, *supra* note 2, at 15 ("If a particular high-impact use case is not compliant with the minimum practices then the agency must safely discontinue use of the AI functionality."); Biden AI Memo, *supra* note 1, at 15 ("Except as prevented by applicable law and governmentwide guidance, agencies must apply the minimum risk management practices in this section to safety-impacting and rights-impacting AI by December 1, 2024, or else stop using the AI until they achieve compliance").

III. THE BIDEN AI MEMO

The Biden AI Memo owes its origins to a decade of advocacy, research, and prior government work, ultimately leading to the Blueprint for an AI Bill of Rights ("AI Bill of Rights"), a white paper released by the Biden Administration's White House Office of Science and Technology Policy (OSTP) in October of 2022.[42] The AI Bill of Rights contains principles and practices designed to protect the public from the potential harms of AI, and the Biden AI Memo drew not only inspiration, but actual substantive provisions directly from it. This Part will detail the origins of the AI Bill of Rights and describe how it influenced the Biden AI Memo.

A. HISTORY OF AI-RELATED PRINCIPLES AND RESEARCH ON AI HARMS

In 2008, Chris Anderson, editor-in-chief of Wired magazine, famously described the advent of "Big Data" as "the end of theory."[43] The declining price of computer technology and the availability of cheap data had remade the economy. The "Big Data" era of big business was in full swing. Government had also gotten on board the data train, with the Obama Administration pushing data as an important asset, but with a focus on open data—data "for the people."[44]

The year 2014 marked a turning point on our collective journey towards understanding data-driven technologies. That year, members of civil society, government practitioners, journalists, and researchers from law, computer science, and other related fields began to seriously examine the societal impacts of Big Data. All these different groups were realizing that Big Data—the then-popular term that has since been overtaken by AI—can do harm as well as good. A coalition of civil rights groups released the "Civil Rights Principles for the Era of Big Data"[45] which advocates for fairness, personal control of information, protection from inaccurate data, and other safeguards from big data related inference. The Obama Administration's OSTP released a

42. WHITE HOUSE OFF. OF SCI. & TECH. POL'Y, BLUEPRINT FOR AN AI BILL OF RIGHTS: MAKING AUTOMATED SYSTEMS WORK FOR THE AMERICAN PEOPLE (Oct. 4, 2022) [hereinafter AI Bill of Rights].

43. Chris Anderson, *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, WIRED (June 3, 2008), https://www.wired.com/2008/06/pb-theory/ [https://perma.cc/E6PE-795Y].

44. *FACT SHEET: Data by the People, for the People*, OBAMA WHITE HOUSE ARCHIVES (Sep. 28, 2016), https://obamawhitehouse.archives.gov/the-press-office/2016/09/28/fact-sheet-data-people-people-eight-years-progress-opening-government [https://perma.cc/85A3-5UP3].

45. THE LEADERSHIP CONFERENCE ON CIVIL AND HUMAN RIGHTS, CIVIL RIGHTS PRINCIPLES FOR THE ERA OF BIG DATA (2014), https://civilrights.org/2014/02/27/civil-rights-principles-era-big-data/ [https://perma.cc/PKF8-HMBA].

landmark report observing that "big data analytics have the potential to eclipse longstanding civil rights protections in how personal information is used in housing, credit, employment, health, education, and the marketplace."[46] Later that year, the first convening of academic researchers focused on the principles of fairness, accountability, and transparency in machine learning (FAT/ML) was held at a computer science conference with interdisciplinary attendees across computer science, law, and policy, including some of the individuals involved in developing the big data civil rights principles.[47]

In the years following, these groups continued working and meeting, and FAT/ML convenings occurred yearly. In 2016, OSTP released a second report on big data, this time focused more squarely on its impact on civil rights related to credit, employment, higher education, and criminal justice.[48] Journalists also released key investigations, including a landmark ProPublica article from 2016 showing racial bias in the incorrect predictions of a pretrial risk assessment.[49] Further work by journalists and researchers identified concerns with the use of AI in domains including predictive policing,[50] online advertising,[51] and many others.[52] Academic interest in the field grew, with over 450 attendees at the

46. EXEC. OFF. OF THE PRESIDENT, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES 3 (2014), https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf [https://perma.cc/AW6H-3Y2N].

47. The speakers page from the 2014 FAT/ML website lists both David Robinson and Harlan Yu, who were at the time the principals of Upturn, a DC-based tech policy nonprofit that was involved in the development of the civil rights principles. *See 2014 Speakers*, FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY IN MACH. LEARNING, https://www.fatml.org/schedule/2014/speakers [https://perma.cc/HN63-G4XQ]; AARON RIEKE, DAVID ROBINSON & HARLAN YU, CIVIL RIGHTS, BIG DATA, AND OUR ALGORITHMIC FUTURE (2014), https://www.upturn.org/work/civil-rights-big-data-and-our-algorithmic-future/ [https://perma.cc/53JM-XGXQ] (indicating that David Robinson and Harlan Yu served as technical advisors for the Big Data and Civil Rights Principles).

48. EXEC. OFF. OF THE PRESIDENT, BIG DATA: A REPORT ON ALGORITHMIC SYSTEMS, OPPORTUNITY, AND CIVIL RIGHTS (2014), https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf [https://perma.cc/T753-2VN3].

49. Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks.*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/24EQ-RNQA].

50. William Isaac & Kristian Lum, *To Predict and Serve?*, 13 SIGNIFICANCE 14 (Oct. 7, 2016), https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00960.x.

51. Julia Angwin & Terry Parris Jr., *Machine Bias: Facebook Lets Advertisers Exclude Users by Race*, PROPUBLICA (Oct. 28, 2016), https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race [https://perma.cc/N8EE-FZ2K].

52. *Machine Bias: Investigating Algorithmic Injustice*, PROPUBLICA, https://www.propublica.org/series/machine-bias [https://perma.cc/P2AR-YYL4].

first full-size conference offering of the Conference on Fairness, Accountability, and Transparency in 2018.[53]

With this growing public and researcher interest came a movement to develop principles and practices for governing AI. In 2016, researchers from the FAT/ML community developed principles and an associated transparency reporting mechanism they called a "social impact statement,"[54] with others' calls for algorithmic impact assessments as a legal intervention following close behind.[55] Companies also began releasing AI ethics principles, including IBM in 2017,[56] Google in 2018,[57] and Microsoft in 2018 (specific to facial recognition).[58] Following these releases, governmental bodies also released AI ethics principles: more than forty countries signed on to the OECD AI Ethics Principles in 2019[59] and the Pentagon and U.S. Intelligence Community

53. Press Release, Conference on Fairness, Accountability, and Transparency (Jan. 30, 2018), https://facctconference.org/2018/press_release.html [https://perma.cc/4HYP-CW26].

54. Nicholas Diakopoulos, Sorelle Friedler, Marcelo Arenas, Solon Barocas, Michael Hay, Bill Howe, HV Jagadish, Kris Unsworth, Arnaud Sahuguet, Suresh Venkatasubramanian, Christo Wilson, Cong Yu, & Bendert Zevenbergen, *Principles for Accountable Algorithms and a Social Impact Statement for Algorithms*, FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY IN MACH. LEARNING, https://www.fatml.org/resources/principles-for-accountable-algorithms [https://perma.cc/8BWZ-YA42] (referencing the Dagstuhl working group write-up from the 2016 Dagstuhl Seminar, *Data, Responsibly*: https://www.dagstuhl.de/16291 [https://perma.cc/UX7M-9847]).

55. Andrew D. Selbst, *Disparate Impact in Big Data Policing*, 52 GA. L. REV. 109, 169–82 (2017); DILLON REISMAN, JASON M. SCHULTZ, KATE CRAWFORD & MEREDITH WHITTAKER, ALGORITHMIC IMPACT ASSESSMENTS REPORT: A PRACTICAL FRAMEWORK FOR PUBLIC AGENCY ACCOUNTABILITY (2018), https://ainowinstitute.org/publication/algorithmic-impact-assessments-report-2 [https://perma.cc/7X9G-3V5A].

56. Contemporary discussion of the principles dates them to 2017. *See* Alison DeNisco Rayome, *3 Guiding Principles for Ethical AI, from IBM CEO Ginni Rometty*, TECHREPUBLIC (Jan. 17, 2017), https://www.techrepublic.com/article/3-guiding-principles-for-ethical-ai-from-ibm-ceo-ginni-rometty/ (IBM principles available at: https://web.archive.org/web/20210416170025/https://www.ibm.com/policy/wp-content/uploads/2018/06/IBM_Principles_SHORT.V4.3.pdf [https://perma.cc/KY4L-X5V3]).

57. Sundar Pichai, *AI at Google: Our Principles*, GOOGLE (June 7, 2018), https://blog.google/technology/ai/ai-principles/ [https://perma.cc/8T5U-ERBD].

58. *Six Principles for Developing and Deploying Facial Recognition Technology*, MICROSOFT, https://msblogs.thesourcemediaassets.com/sites/5/2018/12/MSFT-Principles-on-Facial-Recognition.pdf [https://perma.cc/PGA4-PZX4].

59. *See* ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, RECOMMENDATION OF THE COUNCIL ON ARTIFICIAL INTELLIGENCE (May 2019), https://one.oecd.org/document/C/MIN(2019)3/FINAL/en/pdf [https://perma.cc/Y6FM-WBEQ] (discussing the AI Ethics Principles); ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, RECOMMENDATION OF THE COUNCIL ON ARTIFICIAL INTELLIGENCE: ADHERENTS, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449#

released principles in 2020.[60] During President Trump's first term, EO 13960 included AI principles as well. The principles from all of these groups were largely overlapping. More than half of the eighty-four such documents analyzed in 2019 included the principles of transparency, fairness, non-maleficence, responsibility, and privacy.[61]

B. THE AI BILL OF RIGHTS

The next important step on the way to the Biden AI Memo was a white paper titled "Blueprint for an AI Bill of Rights."[62] Under President Biden, OSTP was tasked with developing equity-oriented technology policy, both related to AI and more generally.[63] The AI Bill of Rights was one outcome of OSTP's work on equity and AI policy. It puts forth five AI principles designed to protect the public from the potential harms of AI when used to impact people's rights, opportunities, or access to critical resources. These principles are: (1) safe and effective systems; (2) algorithmic discrimination protections; (3) data privacy; (4) notice and explanation; and (5) human alternatives, consideration, and fallback.[64]

The developed principles built on the other principles of the era,[65] originating from civil society, academia, and the private sector, and the AI Bill

adherents [https://perma.cc/TR94-54QV] (listing the countries that signed on to the AI Ethics Principles).

60. C. Todd Lopez, *DOD Adopts 5 Principles of Artificial Intelligence Ethics*, U.S. DEP'T OF WAR (Feb. 25, 2020), https://www.defense.gov/News/News-Stories/article/article/2094085/dod-adopts-5-principles-of-artificial-intelligence-ethics/ [https://perma.cc/8ASX-XEPG]; *Principles of Artificial Intelligence Ethics for the Intelligence Community*, DIR. OF NAT'L INTEL., https://www.dni.gov/files/ODNI/documents/Principles_of_AI_Ethics_for_the_Intelligence_Community.pdf (For the historical press release, see: https://web.archive.org/web/20250213213506/https://www.dni.gov/index.php/newsroom/press-releases/press-releases-2020/3468-intelligence-community-releases-artificial-intelligence-principles-and-framework).

61. Anna Jobin, Marcello Ienca & Effy Vayena, *The Global Landscape of AI Ethics Guidelines*, 1 NATURE MACHINE INTELLIGENCE 389, 391–396 (2019), https://www.nature.com/articles/s42256-019-0088-2 [https://perma.cc/Q8J7-DWBL] (*see* especially Table 3, at 395).

62. AI Bill of Rights, *supra* note 42.

63. Letter from Joseph Biden, President-Elect, to Dr. Eric S. Lander, President's Sci. Advisor and Dir. of the Off. of Sci. & Tech. Pol'y (Jan. 15, 2021), https://bidenwhitehouse.archives.gov/ostp/news-updates/2021/01/15/a-letter-to-dr-eric-s-lander-the-presidents-science-advisor-and-director-of-the-office-of-science-and-technology-policy/ [https://perma.cc/RWP7-T5TX].

64. AI Bill of Rights, *supra* note 62, at 5–7.

65. While these five principles are not exactly the same as the five consensus principles identified above (*see supra* note 62 and accompanying text), they do map roughly onto these ones: non-maleficence maps to safe and effective systems, fairness to algorithmic

of Rights was described by its writers as focused on "protecting [our] civil rights in the algorithmic age."[66] The AI Bill of Rights aimed to beyond principles, toward implementation as well. Accordingly, in a section titled "From Principles to Practice: A Technical Companion to the Blueprint for an AI Bill of Rights," it described practices needed to achieve these principles in detail.[67]

In the month following the release of the AI Bill of Rights, OpenAI's ChatGPT launched.[68] The dramatically increased public attention on AI resulted in a flurry of White House actions.[69] In October 2023, President Biden

discrimination protections, privacy to privacy, transparency to notice and explanation, and responsibility to human alternatives, consideration, and fallback.

66. Alondra Nelson, Sorelle Friedler & Ami Fields-Meyer, *Blueprint for an AI Bill of Rights: A Vision for Protecting Our Civil Rights in the Algorithmic Age*, OSTP BLOG (Oct. 4, 2022), https://bidenwhitehouse.archives.gov/ostp/news-updates/2022/10/04/blueprint-for-an-ai-bill-of-rightsa-vision-for-protecting-our-civil-rights-in-the-algorithmic-age/ [https://perma.cc/MYN6-ZNWF].

67. AI Bill of Rights, *supra* note 62, at 18–20, 26–28, 33–38, 43–44, 49–51 (discussing five subsections (one per principle) titled, "What should be expected of automated systems").

68. *Introducing ChatGPT*, OPENAI (Nov. 30, 2022) (OpenAI Product Release).

69. These actions included: an executive order from February 2023 that, in part, worked to prevent algorithmic discrimination; in May 2023, the announcement of investment in research, a public red-teaming challenge, and the future release of the Biden AI Memo later that year for public comment; a National R&D Strategy released in May 2023; a White House announcement of voluntary commitments from companies to manage AI risks in July 2023; an AI Cyber Challenge launched in August 2023; and further voluntary commitments from companies secured in September 2023. *See* Exec. Order No. 14091 on Further Advancing Racial Equity and Support for Underserved Communities Through the Federal Government. 88 Fed. Reg. 10825 (Feb. 16, 2023); *FACT SHEET: Biden-Harris Administration Announces New Actions to Promote Responsible AI Innovation that Protects Americans' Rights and Safety* (May 4, 2023), https://bidenwhitehouse.archives.gov/ostp/news-updates/2023/05/04/fact-sheet-biden-harris-administration-announces-new-actions-to-promote-responsible-ai-innovation-that-protects-americans-rights-and-safety/ [https://perma.cc/BGD3-QYXZ]; SELECT COMM. ON A.I. OF THE NAT'L SCI. & TECH. COUNCIL, NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT STRATEGIC PLAN 2023 UPDATE (May 2023), https://bidenwhitehouse.archives.gov/wp-content/uploads/2023/05/National-Artificial-Intelligence-Research-and-Development-Strategic-Plan-2023-Update.pdf [https://perma.cc/45D8-QNNB]; *FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI*, BIDEN WHITE HOUSE ARCHIVES (July 21, 2023), https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/ [https://perma.cc/5Q4E-46PN]; *Biden-Harris Administration Launches Artificial Intelligence Cyber Challenge to Protect America's Critical Software*, BIDEN WHITE HOUSE ARCHIVES (Aug. 9, 2023), https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2023/08/09/biden-harris-administration-launches-artificial-intelligence-cyber-challenge-to-protect-americas-critical-software/ [https://perma.cc/P7FZ-EZLB]; *FACT SHEET: Biden-Harris*

signed EO 14110 on "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence." EO 14110 was a broad statement of policy regarding both the desire to promote AI innovation and address the risks associated with it.[70] As part of its overall push to address risks to safety and rights, the EO required that OMB issue guidance specifying, among other things:

> required minimum risk-management practices for Government uses of AI that impact people's rights or safety, *including, where appropriate, the following practices derived from OSTP's Blueprint for an AI Bill of Rights* and the NIST AI Risk Management Framework: conducting public consultation; assessing data quality; assessing and mitigating disparate impacts and algorithmic discrimination; providing notice of the use of AI; continuously monitoring and evaluating deployed AI; and granting human consideration and remedies for adverse decisions made using AI.[71]

Thus, OMB was directed to make many of the provisions of the AI Bill of Rights binding.

C. THE AI BILL OF RIGHTS' INFLUENCE ON THE BIDEN AI MEMO

Given the commands to OMB to draw on a variety of sources (EO 13960, the AI Bill of Rights, and the National Institute of Standards and Technology (NIST) AI Risk Management Framework),[72] it was not obvious that the resulting memo would draw so heavily on the AI Bill of Rights in particular. But there is a key perspective that the AI Bill of Rights and the Biden AI Memo share: there are specific identified impacts of AI that are worth protecting against. This perspective stems from the AI Bill of Rights's focus on practical implementation to protect civil rights. In both documents, the specific impacts

---

Administration Secures Voluntary Commitments from Eight Additional Artificial Intelligence Companies to Manage the Risks Posed by AI, BIDEN WHITE HOUSE ARCHIVES (Sep. 12, 2023), https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2023/09/12/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-eight-additional-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/ [https://perma.cc/E3WP-YPLM].

70. EO 14110, *supra* note 12, §§ 1–2.

71. *Id.* § 10(1)(iv) (emphasis added).

72. While EO 14110 played a large role, it was not the sole legal basis for the Biden AI Memo. As discussed above, both the AI in Government Act, *supra* note 10, and the Advancing American AI Act, *supra* note 11, at 3670, § 7224(a)(2) (directing OMB to draw on "the principles articulated in EO 13960") played important roles as well. In fact, the Biden AI Memo was clearly in progress prior to the issuance of EO 14110, since only two days after that executive order Vice President Harris announced the release of the 26-page draft memo for public comment. *See OMB Releases Implementation Guidance Following President Biden's Executive Order on Artificial Intelligence*, BIDEN WHITE HOUSE ARCHIVES (Nov. 1, 2023), https://bidenwhitehouse.archives.gov/omb/briefing-room/2023/11/01/omb-releases-implementation-guidance-following-president-bidens-executive-order-on-artificial-intelligence/ [https://perma.cc/WYP7-4CMW].

of concern are identified by the language scoping what AI use cases require instituted protections. In contrast, the NIST AI Risk Management Framework explicitly rules out taking a stance on what specific impacts are worth focusing on, stating that the framework "does not prescribe risk tolerance," or in other words, that organizations with a high risk tolerance can successfully implement the framework by taking few or no steps to manage risk.[73] Organizations are encouraged by NIST to identify the risks they care about and take steps based on those identified risks. OMB instead identifies for agencies a specific set of impacts with risks that must be managed. This contrast demonstrates that the Biden AI Memo looks the way it does at least in part because it follows the civil rights framing of the AI Bill of Rights.

The specific scoping language identifying AI impacts that require protections also makes clear the influence of the AI Bill of Rights on the Biden AI Memo. For example, both the Biden AI Memo and AI Bill of Rights list specific minimum risk management practices when the government uses AI systems that affect people's rights.[74] In both cases, the definitions apply protections to systems grouped into three buckets based on (1) impact to civil rights, (2) equal opportunities, or (3) access to government services, with only slight differences in language between the two. The Biden AI Memo's category of "rights-impacting AI" applies to AI that affect:

> 1) Civil rights, civil liberties, or privacy, including but not limited to freedom of speech, voting, human autonomy, and protections from discrimination, excessive punishment, and unlawful surveillance;
>
> 2) Equal opportunities, including equitable access to education, housing, insurance, credit, employment, and other programs where civil rights and equal opportunity protections apply; or
>
> 3) Access to or the ability to apply for critical government resources or services, including healthcare, financial services, public housing, social services, transportation, and essential goods and services.[75]

The AI Bill of Rights, in turn, applies to automated systems that affect:

> Civil rights, civil liberties, and privacy, including freedom of speech, voting, and protections from discrimination, excessive punishment,

73. U.S. DEP'T OF COM. NAT'L INST. OF STANDARDS & TECH., NIST AI 100-1: ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0) 7 (Jan. 2023), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf [https://perma.cc/5NV8-8Q5T].

74. The Biden AI Memo focuses on the category of "rights-impacting AI," while the AI Bill of Rights defines its scope as automated systems that affect "rights, opportunities, or access." *Compare* Biden AI Memo, *supra* note 1, at 15, *with* AI Bill of Rights, *supra* note 62, at 8.

75. Biden AI Memo, *supra* note 1, at 29 (emphasized for comparison).

> unlawful surveillance, and violations of privacy and other freedoms in both public and private sector contexts;
>
> Equal opportunities, including equitable access to education, housing, credit, employment, and other programs; or,
>
> Access to critical resources or services, such as healthcare, financial services, safety, social services, non-deceptive information about goods and services, and government benefits.[76]

The extent of coverage is somewhat different between the two documents: the Biden AI Memo requires that the "output serves as a principal basis for a decision"[77] while the AI Bill of Rights applies to systems which "have the potential to meaningfully impact" individuals or communities.[78] But the influence of the coverage categories from the AI Bill of Rights on the Biden AI Memo is clear.

There is also a direct relationship between the specific practices detailed in each document. For example, both documents discuss the important practice of testing an AI system to make sure it will work in its real-world context. The specific descriptions of that practice are different in the two documents, yet some sentences are the same and the paragraph structure is similar. (See Appendix A for a detailed comparison.) For example, both documents say that such testing should ensure that the system "will work in its [intended] real-world context," that the "testing should follow domain-specific best practices, when available," and that it "should take into account both the specific technology used and" the humans who make up part of the overall system (e.g., "human operators" or reviewers).[79]

Organizationally, the documents have different structures. The Biden AI Memo organizes the minimum risk management practices into two main categories: those required for safety-impacting and rights-impacting AI systems, and those required solely for rights-impacting AI systems. The AI Bill of Rights instead organizes the associated practices based on the principle they support, with practices appearing under the subheading "What should be expected of automated systems."[80] Broadly, the practices associated with the Safe and Effective Systems principle of the AI Bill of Rights are reorganized in the Biden AI Memo to be required by both safety and rights-impacting AI systems; the practices associated with Algorithmic Discrimination Protections are required solely by rights-impacting AI systems; and some practices

76. AI Bill of Rights, *supra* note 62, at 8.
77. Biden AI Memo, *supra* note 1, at 15.
78. AI Bill of Rights, *supra* note 62, at 8.
79. Biden AI Memo, *supra* note 1, at 18; AI Bill of Rights, *supra* note 62, at 18.
80. AI Bill of Rights, *supra* note 62, at 18.

supporting Notice and Explanation and Human Alternatives, Consideration, and Fallback are required for both rights-impacting and safety-impacting systems, while others are required solely for rights-impacting systems.[81]

For example, the practice of testing in a real-world context appears under Safe and Effective Systems in the AI Bill of Rights, while it is part of the minimum practices for both safety-impacting and rights-impacting AI in the Biden AI Memo. Similarly, the AI Bill of Rights practice to "[p]rovide timely human consideration and remedy by a fallback and escalation system in the event that an automated system fails, produces error, or you would like to appeal or contest its impacts on you" is listed under the Human Alternatives, Consideration, and Fallback principle,[82] while the same practice becomes a requirement to "[m]aintain human consideration and remedy processes" in the Biden AI Memo, listed as a minimum risk-management practice for rights-impacting AI.[83] The memo further states: "[w]here practicable and consistent with applicable law and governmentwide guidance, agencies must provide *timely human consideration and* potential *remedy*, if appropriate, to the use of the AI via *a fallback and escalation system in the event that* an impacted individual would like to *appeal or contest* the AI's negative impacts on them," with emphasis added to phrases that match those from the AI Bill of Rights.[84]

Overall, while the organizational structure of the AI Bill of Rights and Biden AI Memo are markedly different, the definitions, coverage, and protective practices have a clear throughline.

---

81. The Biden AI Memo does not focus on privacy protections corresponding to the AI Bill of Rights Data Privacy principle. Biden AI Memo, *supra* note 1, at 3 ("This memorandum does not address issues that are present regardless of whether AI is used versus any other software, such as issues with respect to . . . privacy . . . . "). It does discuss privacy alongside other issues with some frequency; it is just not a focus. Anecdotally, this is because some people within the government see the Privacy Act of 1974 as doing all the work necessary to secure privacy, making it difficult to internally justify new privacy rules. Structurally, it may also be because of the organizational structure of OMB, with the OIRA privacy office governing privacy-related concerns, while the Office of the Federal CIO led the OMB AI Memo development. *See Office of Information and Regulatory Affairs*, OBAMA WHITE HOUSE ARCHIVES, https://obamawhitehouse.archives.gov/omb/oira [https://perma.cc/RN6U-5RWR] (discussing inclusion of privacy in OIRA); *see also* AI, CIO COUNCIL, https://www.cio.gov/tags/ai/ (documents from OFCIO describing and analyzing the Biden AI Memo) (on file with the Berkeley Technology Law Journal); *White House Releases New Policies on Federal Agency AI Use and Procurement*, WHITE HOUSE (Apr. 7, 2025), https://www.whitehouse.gov/articles/2025/04/white-house-releases-new-policies-on-federal-agency-ai-use-and-procurement/ [https://perma.cc/3D23-BVPR] (quoting an OFCIO Officer on the release of the Trump AI Memo).

82. AI Bill of Rights, *supra* note 62, at 49.

83. Biden AI Memo, *supra* note 1, at 23.

84. *Id.*

IV. THE TRUMP AI MEMO

Given the shift in priorities from the Biden Administration to the Trump Administration, one would reasonably expect key policy changes relating to government use of AI. Indeed, one of the first executive orders of the second Trump Administration revoked President Biden's executive order on AI (EO 14110) which had (in part) directed OMB to create the Biden AI Memo.[85] An additional executive order directed OMB to revise the Biden AI Memo within sixty days of the order's January 23, 2025 signing date.[86] OMB was directed to draft a new memo consistent with the new Administration's policy on AI: "It is the policy of the United States to sustain and enhance America's global AI dominance in order to promote human flourishing, economic competitiveness, and national security."[87] But while the Trump AI Memo introduced some important changes, what is perhaps more notable is a surprising degree of continuity between the two OMB AI Memos.

Some of the continuity can be traced to specific policy and associated implementation that survived administration changes. The first executive branch efforts to govern internal use of AI stemmed from the first Trump Administration's EO 13960, as discussed above.[88] It instructs the government to adhere to nine principles for government use of AI and directs OMB to create policy to facilitate them.[89] Neither President Biden nor President Trump ever rescinded it. The principles were subsequently incorporated into the legal requirements of the OMB AI Memos under the Advancing American AI Act.[90] To be sure, an important component of the Biden Administration policy on

85. Exec. Order No. 14148 on Initial Rescissions of Harmful Executive Orders and Actions, 90 Fed. Reg. 8237 (2025) (revoking EO 14110 in Section 2(ggg)) [hereinafter EO 14148].

86. Exec. Order No. 14179 on Removing Barriers to American Leadership in Artificial Intelligence, 90 Fed. Reg. 8741 (2025), at Section 5(b). It also directed OMB to revise the associated Biden memo on AI procurement (M-24-18).

87. Exec. Order 14179 on Removing Barriers to American Leadership in Artificial Intelligence, 90 Fed. Reg. 8741 (2025), at Section 2.

88. *See supra* notes 14, 32, and 61, and the accompanying text.

89. The nine principles are: "Lawful and respectful of our Nation's values"; "Purposeful and performance-driven"; "Accurate, reliable, and effective"; "Safe, secure, and resilient"; "Understandable"; "Responsible and traceable"; "Regularly monitored"; "Transparent"; and "Accountable." EO 13960, *supra* note 32. These nine principles do not match the five consensus principles mentioned earlier, see *supra* note 61, yet there are many similarities. Non-maleficence is core to many of these (purposeful and performance-driven; accurate, reliable and effective; safe, secure, and resilient; and regularly monitored), transparency appears directly, and responsibility appears as both responsible and traceable and accountable.

90. *See* Advancing American AI Act, *supra* note 11, at 3669, § 722(4)(a) (describing documents the Director shall consider in issuing the required guidance on government use of AI.).

AI was its focus on equity, and these principles do not mention it (though the principle "Lawful and respectful of our Nation's values" explicitly includes civil rights in the explanatory paragraph).[91] Accordingly, and unsurprisingly, the Trump AI Memo removes any explicit discussion of equity from its replacement memo.[92] EO 13960, by contrast, was hardened by the statute adopting it as well as by the Biden Administration's efforts to implement the executive order by creating the required AI use case inventory, designating officials at each agency responsible for coordinating work on AI, and generally creating plans for compliance with the EO.[93] This continued implementation-work and its legal basis serve as an important, consistent, and principle-based through-line between the first Trump, Biden, and second Trump Administrations.

The core features of the OMB AI Memos have not changed much. The Trump AI Memo retains much of the same structure and requirements as the Biden AI Memo. The definition of AI—and thus the systems that are covered—has not changed.[94] The general structure of the substantive requirements—identifying some AI systems as requiring specific minimum practices based on use case[95] and prohibiting their use if the minimum practices are not met[96]—has not changed. The governance structure—including the

91. EO 13960, *supra* note 32.

92. In addition to these executive orders focused on AI, the Trump Administration also revoked Biden's equity-related executive orders, including EO 14091, which contained the algorithmic discrimination definition used in the Biden AI Memo. *See* EO 14148, *supra* note 85 (revoking EOs 13985 and 14091).

93. As examples of such agency compliance plans, see U.S. DEP'T OF THE TREASURY, OFF. OF THE CHIEF INFO. OFF., U.S. DEPARTMENT OF THE TREASURY EXECUTIVE ORDER 13960 CONSISTENCY PLAN (Dec. 2022), https://home.treasury.gov/system/files/136/Treasury-EO13960-Consistency-Plan.pdf [https://perma.cc/JK7N-XK4G]; *see also* Bhavya Lal & Kate Calvin, *NASA's Responsible AI Plan*, NAT'L AERONAUTICS & SPACE ADMIN. (2022), https://ntrs.nasa.gov/api/citations/20220013471/downloads/RAI%20Plan%20Sept%201%202022.pdf [https://perma.cc/V87P-QWKE]. AI use case inventory guidance was also issued by OFCIO before the Biden AI Memo as part of EO 13960 implementation. *See EO 13960: Artificial Intelligence (AI) Use Case Inventories*, OFCIO (2023), https://www.cio.gov/assets/resources/2023-Guidance-for-AI-Use-Case-Inventories.pdf [https://perma.cc/QT7W-NLJY].

94. *Compare* Trump AI Memo, *supra* note 2, at 18, *with* Biden AI Memo, *supra* note 1, at 26–27.

95. *Compare* Trump AI Memo, *supra* note 2, at 19 (defining "high-impact AI"), *with* Biden AI Memo, *supra* note 1, at 29–30 (defining "rights-impacting AI" and "safety-impacting AI").

96. *Compare* Trump AI Memo, *supra* note 2, at 15 ("If a particular high-impact use case is not compliant with the minimum practices then the agency must safely discontinue use of the AI functionality."), *with* Biden AI Memo, *supra* note 1, at 14 ("[A]gencies must implement the minimum practices in Section 5(c) of this memorandum for safety-impacting and rights-impacting AI, or else stop using any AI in their operations that is not compliant with the minimum practices, consistent with the details and caveats in that section.").

designation of agency CAIOs—has not changed, nor has the requirement for agencies to publicly report their AI use cases.[97]

When comparing the memos' actual texts, we see that the Trump Administration introduced some cuts, but much of the text itself was retained, with some text being moved around. It seems fairly clear that the authors involved aimed to make the new memo have fewer words overall, in addition to substantive policy changes to the memo.[98] As a result, some of the cut language is ambiguous—was the specificity of the Biden memo language cut because it was deemed extraneous, or because of a policy rationale? It can be hard to say.

As we see it, while the continuity is probably most notable, there are also two high-level changes to the Trump AI Memo worth drawing out: (1) a move toward risk regulation and away from a recognition and focus on individualized harms or rights—notably in line with what Professor Margot Kaminski identifies as the convergence of global AI laws around risk regulation[99]—and (2) a deprioritization of equity, bias, or discrimination harm in line with the Trump Administration's general hostility to such ideas.[100]

The move toward risk regulation can be seen in a couple differences. The most notable is the category changes. As discussed in Part I, the core use-based definition was changed from a split "safety-impacting AI" and "rights-impacting AI" to a single "high-impact AI" category that largely combines the previous two categories, with a single set of minimum risk management practices now being required. This approach mirrors other risk regulation regimes, such as the EU's AI Act, which, while concerned with both safety

97. This was a requirement instituted by EO 13960 issued under the first Trump Administration and by the Advancing American AI Act, *supra* note 11, at 3672, § 7225(a)(3).

98. The Biden AI Memo is 34 pages and approximately 14,700 words while the Trump AI Memo is 25 pages and approximately 9,700 words. Reducing the number of words as a goal in and of itself has been a focus of the Trump Administration in other settings. *See* David Gilbert & Vittoria Elliott, *DOGE Put a College Student in Charge of Using AI to Rewrite Regulations*, WIRED (Apr. 30, 2025), https://www.wired.com/story/doge-college-student-ai-rewrite-regulations-deregulation/ [https://perma.cc/DLA5-CKZF].

99. Margot E. Kaminski, *Regulating the Risks of AI*, 103 B.U. L. REV. 1347, 1351 (2023).

100. *See, e.g.*, Adam Serwer, *The Great Resegregation*, ATLANTIC (Feb. 22, 2025), https://www.theatlantic.com/politics/archive/2025/02/trump-attacks-dei/681772/ [https://perma.cc/U3YD-Z2HE] (observing that The Trump Administration aims to "reverse the civil-rights movement").

and fundamental rights,[101] packs both concepts into a regulation of "high risk" AI.[102]

Other aspects of this move towards a risk regulation approach include the removal of two requirements: that government agencies "[n]otify negatively affected individuals"[103] and that the agencies "[m]aintain options to opt-out for AI-enabled decisions."[104] Both memos contain minimum risk management practices, including pre-deployment testing, AI impact assessments (including independent review), ongoing monitoring, human training, remedy and appeals processes, and public consultation. But the targeted removal of notice and opt-out rights for individuals seems to indicate a shift away from any individualized right or redress component of the governance regime.

The second change is a move away from equity and discrimination concerns. Without the background understanding of the Trump Administration's general hostility to such concerns, these changes in the memo might be more difficult to interpret. With the collapse of the safety-impacting and rights-impacting categories, there was a fair amount of text that could be considered redundant and could have been deleted for that reason. For example, in the Biden AI Memo, there were two instances of commands to "conduct ongoing monitoring." One appears under the heading "Minimum Practices for Either Safety-Impacting or Rights-Impacting AI."[105] The other is more specific; it commands agencies to "[c]onduct ongoing monitoring and mitigation for AI-enabled discrimination," specifically stating that "[a]s part of the ongoing monitoring requirement [referenced above], agencies must also monitor rights-impacting AI to specifically assess and mitigate AI-enabled discrimination."[106] As this command already appears under the heading "Additional Minimum Practices for Rights-Impacting AI," one could see it as redundant where the categories have been collapsed. Indeed the corresponding language in the Trump AI Memo is a command to "conduct testing and periodic human review of AI use cases, where feasible, to identify any adverse impacts to the performance and security of AI functionality, including those that may violate laws governing privacy, civil rights, or civil

101. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024, art. 1, 2024 O.J. (L 1689) 1 [hereinafter AI Act] ("The purpose of this Regulation is to . . . ensur[e] a high level of protection of health, safety, fundamental rights enshrined in the Charter . . . . ").

102. AI Act, *supra* note 101, art. 6; *see also* Margot Kaminski & Andrew Selbst, *An American's Guide to the EU's AI Act*, 40 BERKELEY TECH. L.J. 1081 (2025).

103. Biden AI Memo, *supra* note 1, at 23.

104. *Id.* at 24.

105. *Id.* at 19.

106. *Id.* at 23.

liberties."[107] If one were to read the concern with discrimination as limited to that which is already illegal, then it is possible to read this change as a non-substantive language cleanup, given the category shift. In a similar vein, though the Biden AI Memo specifically barred the use of noncompliant AI generally, it reiterated that prohibition within the section on discriminatory AI.[108] The Trump AI Memo removes this language, but it is not clear whether this truncation is due to perceived redundancy or meant to be substantive.

There are, however, other clues indicating that the truncations are meant as a substantive change, narrowing the scope of discrimination protections. One clue is the complete removal of the words equity and bias from the text of the memo. Another is the decision to treat the problem of AI bias as limited to already-illegal discrimination. Language around discrimination in the Biden AI Memo directed agencies to "[m]itigate disparities that lead to, or perpetuate, unlawful discrimination *or harmful bias, or that decrease equity* as a result of the government's use of the AI."[109] Thus, the Biden AI Memo did not merely limit the AI bias concerns to that which was already illegal; rather, it focused on mitigating real harms, whether otherwise proscribed or not. The corresponding Trump language on mitigation is combined with the above quoted language on conducting testing, stating simply that "[a]gencies must implement appropriate mitigations" later in the paragraph.[110] Again, this rephrasing is arguably a more efficient use of language, but it is more likely an intentional weakening of the Biden AI Memo's discrimination protections. The Trump Administration could not entirely ignore or remove any discussion of civil rights or civil liberties because these concerns were written into the AI in Government Act[111] and EO 13960,[112] but such a vague command allows individual agency personnel to argue that they have mitigated appropriately while soft-pedaling or stonewalling substantive mitigation.

Notably, the move to narrow bias concerns to only those already proscribed by law is made worse by the Trump Administration's public position on what counts as illegal discrimination. President Trump stated in EO 14281 that not only will the Administration no longer recognize disparate

107. Trump AI Memo, *supra* note 2, at 17.

108. Biden AI Memo, *supra* note 1, at 21 ("Mitigate disparities that lead to, or perpetuate, unlawful discrimination or harmful bias, or that decrease equity as a result of the government's use of the AI; and 4. Consistent with applicable law, cease use of the AI for agency decisionmaking if the agency is unable to adequately mitigate any associated risk of unlawful discrimination against protected classes."); *Id.* at 23 ("Where sufficient mitigation is not possible, agencies must safely discontinue use of the AI functionality.").

109. Biden AI Memo, *supra* note 1, at 17 (emphasis added).

110. Trump AI Memo, *supra* note 2, at 17.

111. AI in Government Act, *supra* note 10, at 2288, § 104(a)(2).

112. EO 13960, *supra* note 32.

impact as a viable theory of discrimination, but that the doctrine "violates our Constitution."[113] This position is obviously not a correct statement of the law as it stands today, but its adoption as the official government position is certainly concerning, especially as disparate impact is a much more important theory for algorithmic discrimination than disparate treatment.[114]

Beyond these two shifts, the intent behind some other changes are more ambiguous. For example, on impact assessments, the Biden AI Memo directs agencies conducting required AI impact assessments to include documentation of:

> 1. *The intended purpose for the AI and its expected benefit*, supported by specific metrics or qualitative analysis. Metrics should be quantifiable measures of positive outcomes for the agency's mission—for example to reduce costs, wait time for customers, or risk to human life—that can be measured using performance measurement or program evaluation methods after the AI is deployed to demonstrate the value of using AI. Where quantification is not feasible, qualitative analysis should demonstrate an expected positive outcome, such as for improvements to customer experience, and it should demonstrate that AI is better suited to accomplish the relevant task as compared to alternative strategies.[115]

The Trump AI Memo also requires AI impact assessments to document the intended purpose and expected benefit of the AI system, described as:

> A. *the intended purpose for the AI and its expected benefit*, supported by specific metrics or qualitative analysis, assessing impact inclusive of but not limited to costs, customer experience, or expected positive outcomes of AI use, as compared to existing agency processes;[116]

Some detail has clearly been eliminated in the Trump AI Memo, yet it is hard to determine the degree to which the elimination of such detail was intended to effect substantive changes to agency practices or whether substantive changes will result. One rationale for interpreting the Trump AI Memo's more parsimonious language as a non-substantive change is that, as the first such management memorandum detailing how agencies should assess their uses of AI, the Biden AI Memo also served an educational and communicative purpose to the agencies, which was no longer necessary in the Trump version. A contrary interpretation is also reasonable. If the Trump Administration

113. Exec. Order No. 14281 on Restoring Equality of Opportunity and Meritocracy, 90 Fed. Reg. 17537 (2025).

114. *See, e.g.*, Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671 (2016).

115. Biden AI Memo, *supra* note 1, at 17.

116. Trump AI Memo, *supra* note 2, at 16.

disagreed with the importance of only using AI where it is clearly superior to alternative processes, removing phrasing and details such as "it should demonstrate that AI is better suited to accomplish the relevant task as compared to alternative strategies" could serve to de-emphasize the requirement.[117] Other changes throughout the Trump AI Memo lead to similar questions.

In the Trump AI Memo, we see remarkable structural continuity from the Biden AI Memo, key substantive changes, and some changes with unclear purposes.

V. FROM PRINCIPLES TO PRACTICE: UNDERSTANDING THE AI MEMOS AS INTERMEDIATE INSTRUMENTS

Overall, while we might have expected a complete overhaul of the governance framework laid out in the Biden AI Memo given the different policy priorities across administrations, the Trump AI Memo contained a surprising amount of continuity, despite the substantive changes we identified above. In this Part, we argue that there are structural reasons for this continuity,[118] and these reasons can help us understand the ultimate function

117. We note that the importance of comparing AI to alternative (likely non-technical) strategies, and not just to other AI systems, has long been argued to be a key aspect of the protections provided by algorithmic impact assessments. *See, e.g.*, Selbst, *supra* note 55, at 176.

118. Some of this continuity can surely also be explained by continuity in personnel and the fact that writing and providing for public comment on the original memo (as required by the AI in Government Act) was a lot of work that they would not want to be required to repeat. Two specific members of the second Trump Administration were also key players in earlier administrations: Lynne Parker, Principal Deputy Director of OSTP under the Second Trump Administration also served for a time in the Biden Administration as the Director of the National Artificial Intelligence Initiative Office in OSTP and in the first Trump Administration as the Assistant Director for Artificial Intelligence in OSTP, and Michael Krastios serves as Director of OSTP in the second Trump and CTO of the first Trump Administration. *See* Press Release, White House, White House Releases New Policies on Federal Agency AI Use and Procurement, https://www.whitehouse.gov/articles/2025/04/white-house-releases-new-policies-on-federal-agency-ai-use-and-procurement/ [https://perma.cc/3D23-BVPR] (quoting Parker describing the Trump AI Memo); Lynne Parker, *OSTP's Continuing Work on AI Technology and Uses That Can Benefit Us All*, BIDEN WHITE HOUSE ARCHIVES, https://bidenwhitehouse.archives.gov/ostp/news-updates/2022/02/03/ostps-continuing-work-on-ai-technology-and-uses-that-can-benefit-us-all/ [https://perma.cc/GB6J-5GWV] (coauthoring a blog post describing the AI work of the Biden Administration, including the AI Bill of Rights); 2016–2019 PROGRESS REPORT: ADVANCING ARTIFICIAL INTELLIGENCE R&D (2019), https://trumpwhitehouse.archives.gov/wp-content/uploads/2019/11/AI-Research-and-Development-Progress-Report-2016-2019.pdf [https://perma.cc/M4XT-2K9K] (mentioning Parker as Co-chair of the Subcommittee on Machine Learning and Artificial Intelligence in the first Trump Administration); Letter from Donald J. Trump, President, to Michael Kratsios, Dir. of the White House Off. of Sci. & Tech. Pol'y (Mar. 26,

of the OMB AI Memos to AI governance, both within the government itself and within the broader AI governance landscape. The OMB AI Memos are therefore best understood as intermediate policy documents, necessary to convert high-level principles into agency action.

As we have argued, the OMB AI Memos are derived from principles aiming to explicate key policy desires in AI governance. The memos are trying to accomplish two different goals. The first is to achieve consistency in implementation of AI principles across the government. The second is to enable practical action across many agencies with highly varied missions and associated AI use cases, but without writing out a separate policy for each individual mission and associated use case—an ill-advised and practically impossible task.

The first goal is accomplished via two different methods: spelling out concrete actions that agencies must take and requiring communication across agencies. The concrete actions are all the substantive requirements discussed above: the creation of a governance framework and especially the CAIO position, reporting of AI use cases, government-wide understanding of which use cases are high-impact, and minimum standards for AI risk mitigation. We call attention to them again here to point out that the very concreteness of the tasks is largely what necessitates such an instrument in the first place. Without such guidance, each agency head would be left to wonder how exactly to cash out high-level principles into organizational changes, policy changes, and reporting policies.

Communication across agencies is accomplished via an interagency council.[119] The interagency council is charged with "coordinat[ing] the development and use of AI across agencies' programs and operations, including enabling compliance with implementation of this memorandum and all other applicable authorities" and on a more practical note, "develop[ing]

2025), https://www.whitehouse.gov/briefings-statements/2025/03/a-letter-to-michael-kratsios-director-of-the-white-house-office-of-science-and-technology-policy/ [https://perma.cc/JY32-EJ4F] (open letter from President Trump charging Kratsios with AI policy in the second Trump Administration); *Michael Kratsios, Chief Technology Officer of the United States*, TRUMP WHITE HOUSE ARCHIVES, https://trumpwhitehouse.archives.gov/people/michael-kratsios/ [https://perma.cc/ZBT4-TZS3] (Kratsios's bio page from the first Trump Administration identifying him as CTO). Nonetheless, we believe the structural reasons we identify are important drivers of the continuity as well.

119. Biden AI Memo, *supra* note 1, at 13–14; Trump AI Memo, *supra* note 2, at 13. The Trump AI Memo specifies that the interagency council is established to "advance the implementation of the AI Principles." Trump AI Memo, *supra* note 2, at 13. The Biden AI Memo does not explicitly include this language.

and promot[ing] shared templates, formats, technical resources, and exemplary uses of agency AI adoption and implementation."[120]

The second goal is enabling the agencies to take practical action on AI risk management without specifying individualized guidance for each AI use case. The agencies have a wide variety of missions and associated AI use cases. For example, the Department of the Interior uses AI to detect invasive bullfrogs,[121] the Department of Veterans Affairs uses AI to create veteran suicide risk assessments,[122] and the Federal Emergency Management Agency uses chatbots to streamline information access internally.[123] As of the 2024 consolidated AI use case inventory, more than two thousand such AI use cases were reported across the federal government.[124] Creating individualized guidance for each such use case would not be feasible or advisable.

This problem is inherent to the need for what we are calling an "intermediate instrument"; it must enable differing implementations of high-level principles. But there are different possible approaches to solving this problem. The approach that the OMB AI Memos took was to create specific requirements and a reporting structure that leave a large amount of room for agency discretion and give deference to agency expertise. There are a number of reasons why this deference may be preferred. It would be logistically infeasible for OMB's single technical office (OFCIO) to oversee the thousands of government AI use cases. Deferring to agencies also ensures that domain experts with relevant knowledge regarding the context of deployment of an AI system are directly involved in its governance[125] (for example ensuring that healthcare professionals in U.S. Department of Health and Human Services oversee medical AI that may require medical expertise to understand and assess). Thus, OMB's guidance is written at an intermediate and cross-sector level that will make sense and be actionable across a wide variety of agencies

120. Trump AI Memo, *supra* note 2, at 13; *accord* Biden AI Memo, *supra* note 1, at 13–14.

121. U.S. DEP'T OF THE INTERIOR, ARTIFICIAL INTELLIGENCE (AI) USE CASE INVENTORY, *supra* note 7.

122. *AI Use Case Inventory*, U.S. DEP'T OF VETERANS AFFS., https://department.va.gov/ai/ai-use-case-inventory/ [https://perma.cc/29PX-3MZP].

123. U.S. DEP'T OF HOMELAND SEC., ARTIFICIAL INTELLIGENCE USE CASE INVENTORY LIBRARY, https://www.dhs.gov/publication/ai-use-case-inventory-library [https://perma.cc/7777-XJSY].

124. *The Government Is Using AI to Better Serve the Public*, AI.GOV, https://web.archive.org/web/20250116141511/https://ai.gov/ai-use-cases/; *see also* Off. of the Fed. Chief Info. Officer, *2024 Federal Agency AI Use Case Inventory*, GITHUB, https://github.com/ombegov/2024-Federal-AI-Use-Case-Inventory/tree/main/data [https://perma.cc/8KR6-MHMK].

125. *See generally* Ryan Calo & Danielle K. Citron, *The Automated Administrative State: A Crisis of Legitimacy*, 70 EMORY L.J. 797 (2021), https://scholarlycommons.law.emory.edu/elj/vol70/iss4/1 [https://perma.cc/EL8S-KWUG].

while satisfying the need for concrete implementation guidance. Such guidance must also include steps that can be straightforwardly checked for compliance by OMB, an office that is unlikely to include domain expertise in the wide variety of AI applications used by agencies and may also not have staff with deep AI expertise. These various limiting factors necessarily lead to concrete requirements that are primarily about the implementation of risk management techniques, rather than focused on substantive outcomes.

Another possible version of an intermediate instrument would have been something akin to Canada's Algorithmic Impact Assessment (AIA),[126] which instructs agencies seeking to use AI to fill out a questionnaire that is turned into an overall risk score.[127] Some of the questions focus on use case (e.g., whether the stakes are "very high"), some on the context (e.g., whether the system is related to health, economic interests, social assistance, access and mobility, licensing and issuance of permits, employment, or other), some on the system capabilities (e.g., image and object recognition or text and speech analysis), and some on risks (e.g., drop-down menus to choose the degree of reversibility of a decision or the degree of risk to rights and freedoms).[128] Many of the questions are accompanied by text boxes for explanations, but they do not factor into the risk score.[129] Finally, there are two pages of mitigation questions that can lower the overall risk score.[130] The risk score then classifies the AI into one of several risk tiers, leading to an escalating set of procedural requirements related to peer review, inequality, notice, human oversight, explainability, training and documentation, IT and business community management, and the internal approval required for the system.[131]

Canada's AIA is a similar intermediate instrument, in that the requirements are written generically in order to avoid writing many more granular assessment instruments across all agencies. But instead of focusing on developing governance and risk mitigation within the agency, the governance mandates exist in this central instrument, with the only solid barrier to

126. *Algorithmic Impact Assessment Tool*, GOV'T OF CANADA, https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html [https://perma.cc/UWL9-DW83].

127. *Directive on Automated Decision-Making*, TREASURY BD. OF CANADA SECRETARIAT (2019), https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592 [https://perma.cc/725H-VCWG].

128. *Algorithmic Impact Assessment Tool*, GOV'T OF CANADA, *supra* note 126.

129. *Id.*

130. *Id.*

131. *Directive on Automated Decision-Making*, TREASURY BD. OF CANADA SECRETARIAT, *supra* note 127, App. C-Impact Level Requirements.

deployment being who must sign off.[132] Unlike a suite of mandatory requirements, the risk scores invite a tradeoff between different types of risk or mitigation to achieve a lower overall tier of risk. This likely trades off simplicity of completing the risk assessment—perhaps thereby increasing willing buy-in—against consistency in requirements. Also, rather than set up lines of communication in which agencies are communicating directly, coupled by a use case inventory, it requires publication of the AIAs directly, such that anyone, including other agencies, can see for themselves what their counterparts are doing.

Our interest here is not in the relative merit of one approach or the other. Rather, it is to point out that what is essential about the OMB AI Memos is precisely their role as intermediate instruments. Wherever there are high-level principles that apply to many different actors in different situations (here, government agencies, but not always), then intermediate instruments like this must exist. Intermediate instruments that, like the OMB AI memos, take the approach of instituting and governing risk management procedures include those from Microsoft,[133] the Government Accountability Office (GAO),[134] and the Data and Trusted AI Alliance,[135] with other companies developing such instruments as products themselves.[136]

Once we understand the role of these intermediate instruments generally, it is easy to see that strategic discussion about the approaches occur at a different level from discussions of the specific policies implemented. Rather, it is about how to reconcile the two goals of consistency in implementation and enabling integration into entities with vastly different missions. A debate over the OMB approach versus the Canadian approach is a different debate than whether, say, equity is a driving value, as important as the second debate

132. *Id.* For Level I and II risk, the assistant deputy minister responsible for the program must provide approval, for Level III it's the head of the department using the AI and for level IV risk, it is the Treasury Board—the Canadian counterpart to OMB. *Id.*

133. *Microsoft Responsible AI Standard, v2 General Requirements*, MICROSOFT (June 2022), https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/final/en-us/microsoft-brand/documents/Microsoft-Responsible-AI-Standard-General-Requirements.pdf?culture=en-us&country=us [https://perma.cc/26T6-3JQQ].

134. GOVERNMENT ACCOUNTABILITY OFFICE, GAO-21-519SP: ARTIFICIAL INTELLIGENCE: AN ACCOUNTABILITY FRAMEWORK FOR FEDERAL AGENCIES AND OTHER ENTITIES (June 2021), https://www.gao.gov/assets/gao-21-519sp.pdf [https://perma.cc/RV3R-HUZ7].

135. *Algorithmic Bias Safeguards*, DATA & TRUST ALL. (July 9, 2024), https://dataandtrustalliance.org/work/algorithmic-safety-mitigating-bias-in-workforce-decisions [https://perma.cc/8MDD-JPWX].

136. *See Scale Trusted AI with Watsonx.Governance*, IBM, https://www.ibm.com/products/watsonx-governance [https://perma.cc/SL8C-YTNM]; *see also AI Governance*, DATAROBOT, https://www.datarobot.com/product/ai-governance/ [https://perma.cc/Q3MJ-A5V4].

separately is. Thus, the central choice that the Biden AI Memo made was to answer the question of how to get from principles to practice, and the degree of continuity between the memos suggests that at least this strategic choice has avoided becoming a partisan battle so far.

VI. CONCLUSION: THE LASTING LEGACY OF THE BIDEN AI MEMO

The OMB AI Memos were designed to give government agencies guidance on how to implement high level AI principles in practice. Both the Biden AI Memo and Trump AI Memo required agencies to implement minimum risk management practices or cease the use of any AI system that could not meet the requirements. As might be expected, when President Trump took over from President Biden, the Administration's priorities shifted drastically, and he immediately moved to counteract much of the work of the Biden Administration. Yet the Biden AI Memo leaves an important and lasting legacy. It lives on both within the government through the structure and strategic choices of the Trump AI Memo and outside the government as a model instrument for implementing AI principles.

Intermediate instruments for AI governance are necessary for two key goals: consistency in implementation of AI principles and practical adoption of AI across a wide variety of use cases. In order to meet these goals, the OMB AI Memos describe practices that agencies must take at a high level of specificity and with deference to agency implementation. This strategic choice, coupled with oversight from OMB, allows a wide variety of agency missions to be accommodated.

Deference to agencies is not necessarily a panacea; agencies' degree of substantive compliance will vary. We have seen such variation already in the outcomes of the Biden AI Memo. For example, the agencies were required to create their AI Use Case Inventories with a deadline that occurred after the 2024 presidential election. Compliance was mixed. The VA's reporting instances included 229 AI use cases, 145 of which were safety or rights-impacting, and almost all had approved extensions. At the same time, the implemented AI systems without extensions had extensive and thoughtful evaluations completed, assessing risks of the use case and connecting those risks explicitly to the concerns about potential violations of AI principles.[137] In

137. As one illustrative example, consider the description of key risks for the REACH VET model that predicts veteran suicide risks:

> Our main concerns regarding risks are as follows: (1) Providers might over-rely on statistical risk estimate. To reduce this risk, we provide information regarding which risk factors were identified by the predictive model to

contrast, the DOJ, which reported 240 AI use cases, 140 of which were safety- or rights-impacting, took no extensions, but reported "None" as the key risks for every implemented use case they were required to assess.[138] This included

> encourage transparent understanding of the basis for the model estimate. Moreover, the clinical prevention program provides only general guidelines to clinicians indicating that identified patients should be reviewed. This forces clinicians to rely on their clinical skills to plan intervention responses. (2) The program could be clinically inefficient and use provider time that might be more effectively focused elsewhere. This is a real risk which may evolve over time as other components of the overall VA suicide prevention program improve, and potentially reduce the need for this targeted prevention program. VHA conducts ongoing evaluations to assess for potential inefficiencies in the model and test methods to improve clinical efficiency. For example, VA is currently piloting a process to expand focus of the REACH VET program to patients not currently engaged in mental health care. (3) The program could lead to unfair allocation of clinical resources if the model is biased and other clinical programming does not compensate for areas of model underperformance. This is also a risk and thus VHA evaluates model bias across key demographic populations and actively develops methods to reduce model bias or address bias by adjusting other components of the treatment system to compensate. After review of the common risks template, the following additional risks were identified: Fair & Equitable (FE)-Algorithm target not reflective of real-world outcome of interest Response: There is a minor risk that the county coroners who assign cause of death might have miscategorized some Veteran deaths, reducing the validity of the outcome variable used. Transparent & Explainable (TE)-Degradation of End-User Trust and Ineffective Challenge or Remedy Processes Response: This is a risk. There has been public misunderstanding of how the model performs and is used. VA has responded and clarified misconceptions in the public forums in which misunderstanding has arisen. Accountable & Monitored (AM)-Performance Efficacy and Fairness (e.g., Model/Data Drift, Model Degradation, or inappropriate application) Response: This is a risk, model performance is degrading over time and an update is in process. However all identified patients are clinically complex and appropriate for clinical attention. Accountable & Monitored (AM)-User-Introduced Errors Response: There is a low risk, there is a community of practice to support proper implementation of the REACH VET Program, which is supervised and trained by national program leads. Accountable & Monitored (AM)-System Performance Not as Intended Response: There is a low risk, the internal VA team that manages the model and risk estimates based on it has validation processes in place to review all steps of the process.

Office of the Federal Chief Information Officer, *2024 Federal Agency AI Use Case Inventory*, GITHUB, https://github.com/ombegov/2024-Federal-AI-Use-Case-Inventory/ (select "data" on the landing page; then select "2024_consolidated_ai_inventory_raw_v2.xls" from the listed files; then click on the "download" button to download and open the file; then scroll to cell number AV1632).

138. *See id.*

ShotSpotter—an AI-based gunshot detection system that includes well-known risks such as over-alerting and inaccuracy[139]—and biometric systems, which are known to have bias risks.[140] Understanding that varied compliance is an inherent feature of deference to the agencies, then, the effectiveness of the OMB AI Memos will depend on the willingness of each administration to enforce its own rules, as well as civil society pressure and public attention in helping to focus the administration on these issues. Thus, while the choices made in the Biden AI Memo make sense in an administration focused on appropriate AI use, they may have kicked the can to future administrations on oversight more than intended. Most importantly for our purposes though, the Biden AI Memo's choice to implement the principles this way is what lives on in the Trump AI Memo.

Outside the government, the OMB AI Memos can also be a useful model of substantive technical guidance where such intermediate instruments are needed. State and local governments share this governance problem. Some have already taken steps to create such governance structures for AI. For example, the Maryland state government has codified a requirement to create such a governance document for state use of AI.[141] Similarly, large companies that operate across different domains and have an interest in implementing of their own AI ethics principles with broad brand consistency across those domains, will also require an intermediate instrument for the same reasons as the government.[142]

---

139. This includes risks raised in a DOJ-funded report. ERIC L. PIZA, GEORGE O. MOHLER, JEREMY G. CARTER, DAVID N. HATTEN, NATHAN T. CONNEALY, RACHAEL ARIETTI, JISOO CHO & EMILY CASTILLO, NAT'L CRIM. JUST. REFERENCE SERV., THE IMPACT OF GUNSHOT DETECTION TECHNOLOGY ON GUN VIOLENCE IN KANSAS CITY AND CHICAGO: A MULTI-PRONGED EVALUATION (2024), https://www.ojp.gov/pdffiles1/nij/grants/308357.pdf [https://perma.cc/3FLU-RDZX]; *see also* a report from the Chicago Police Department OIG. CITY OF CHI., OFF. OF INSPECTOR GEN., THE CHICAGO POLICE DEPARTMENT'S USE OF SHOTSPOTTER TECHNOLOGY (2021), https://igchicago.org/wp-content/uploads/2021/08/Chicago-Police-Departments-Use-of-ShotSpotter-Technology.pdf [https://perma.cc/JAV3-JXJH].

140. Such bias risks, especially with regards to inaccurate facial recognition matches, have been well-documented, including via a governmental report from NIST. *NIST Study Evaluates Effects of Race, Age, Sex on Face Recognition Software*, NAT'L INST. OF STANDARDS & TECH. (Dec. 19, 2019), https://www.nist.gov/news-events/news/2019/12/nist-study-evaluates-effects-race-age-sex-face-recognition-software [https://perma.cc/4KLN-XTV5].

141. MD. CODE, STATE FIN. & PROC. § 3.5-804.

142. For example, Microsoft has both AI Ethics Principles and an accompanying implementation document. *See Microsoft Responsible AI: Principles and Approach*, MICROSOFT, https://www.microsoft.com/en-us/ai/principles-and-approach#ai-principles [https://perma.cc/9QU7-78PR]; *Microsoft Responsible AI Standard, v2 General Requirements*, MICROSOFT, *supra* note 133.

While other models exist, the OMB AI Memos are a particularly useful model of intermediate instruments for AI governance. Some companies create such instruments for internal use and others sell them as governance products.[143] Each likely has their own benefits and drawbacks depending on the use case. Notably, however, the OMB AI Memos are unique in that the government created them in the public interest, without a profit motive. The instruments thus also have greater democratic legitimacy; they are instruments created with a public input process, largely kept in place across both parties' administrations.

After more than a decade of work across civil society, government, and academia to understand AI-driven harms and how to address them, the focus of policy has progressed from analysis to implementation. The OMB AI Memos and similar intermediate instruments are crucial components of this new phase of AI governance, and they should be broadly understood as such.

143. *See, e.g.*, *Scale Trusted AI with Watsonx.Governance*, IBM, *supra* note 136; *see also AI Governance*, DATAROBOT, https://www.datarobot.com/product/ai-governance/ [https://perma.cc/Q3MJ-A5V4].

VII. APPENDIX A

Table 1: Descriptions of Testing Practices as Found in the Biden AI Memo (left) and AI Bill of Rights (right), with Matching Text Highlighted

Biden AI Memo	AI Bill of Rights
Test the AI for performance in a real-world context. Agencies must conduct adequate testing to ensure the AI, as well as components that rely on it, will work in its intended real-world context. Such testing should follow domain-specific best practices, when available, and should take into account both the specific technology used and feedback from human operators, reviewers, employees, and customers who use the service or are impacted by the system's outcomes. Testing conditions should mirror as closely as possible the conditions in which the AI will be deployed. Through test results, agencies should demonstrate that the AI will achieve its expected benefits and that associated risks will be sufficiently mitigated, or else the agency should not use the AI. In conducting such testing, if an agency does not have access to the underlying source code, models, or data, the agency must use alternative test methodologies, such as querying the AI service and observing the outputs or providing evaluation data to the vendor and obtaining results. Agencies are also encouraged to leverage pilots and limited releases, with strong monitoring, evaluation, and safeguards in place, to carry out the final stages of testing before a wider release.	**Testing.** Systems should undergo extensive testing before deployment. This testing should follow domain-specific best practices, when available, for ensuring the technology will work in its real-world context. Such testing should take into account both the specific technology used and the roles of any human operators or reviewers who impact system outcomes or effectiveness; testing should include both automated systems testing and human-led (manual) testing. Testing conditions should mirror as closely as possible the conditions in which the system will be deployed, and new testing may be required for each deployment to account for material differences in conditions from one deployment to another. Following testing, system performance should be compared with the in-place, potentially human-driven, status quo procedures, with existing human performance considered as a performance baseline for the algorithm to meet pre-deployment, and as a lifecycle minimum performance standard. Decision possibilities resulting from performance testing should include the possibility of not deploying the system.

TAKING STANDARDS SERIOUSLY: THE CASE FOR A PRIVATE STANDARDS-BASED APPROACH TO AI GOVERNANCE

Alexander R. Mueller[†] *and Christopher S. Yoo*[††]

ABSTRACT

The rapid proliferation of artificial intelligence (AI) across industry verticals and other domains of social and economic life is forcing policymakers to confront an urgent challenge: how to govern AI systems effectively without stifling innovation or beneficial deployment. This Article argues that private standards-based governance—the adoption of voluntary consensus standards developed through open, multistakeholder processes—offers the most promising "second-best" approach in a world where practical constraints make ideal governance unattainable. Drawing from the successes of standards in governing past digital technologies, the Article highlights how private standards can serve as a regulatory modality for AI by embedding expectations and constraints directly into technological design and organizational practice. Through comparative institutional analysis, it demonstrates how standards-based governance outperforms traditional regulation across four key dimensions: governance architecture, technical expertise and inclusive participation, adaptability to rapid change, and global scalability. At the same time, the Article confronts a number of tradeoffs and challenges such as the nonbinding nature of standards and their susceptibility to industry capture. These challenges and tradeoffs, however, can be managed more efficiently through thoughtful institutional design and strategic government support. The Article ultimately contends that, although standards are not a cure-all, they represent a vital opportunity to build a sustainable, responsive, and highly effective AI governance ecosystem—provided that stakeholders approach their development with the urgency and intentionality the moment demands.

TABLE OF CONTENTS

DOI: https://doi.org/10.15779/Z38WS8HP0D

† Research Fellow, Center for Technology, Innovation & Competition, University of Pennsylvania Carey Law School.

†† Imasogie Professor in Law & Technology, Communications, and Computer & Information Science and Founding Director of the Center for Technology, Innovation & Competition, University of Pennsylvania. This work benefited from discussions following its initial presentation at the 28th Annual BTLJ-BCLT Spring Symposium on "AI Governance at the Crossroads" and was supported by the National Research Foundation of Korea (2022R1A5A7083908).

I. INTRODUCTION

As artificial intelligence (AI) systems grow increasingly capable and widespread, the question of how to govern them effectively has become one of the most vexing contemporary policy challenges. The stakes could hardly be higher: AI promises transformative benefits across healthcare, scientific discovery, and many other domains of human endeavor,[1] yet it also introduces

1. *See, e.g.*, W. Nicholson Price II, *Artificial Intelligence in the Medical System: Four Roles for Potential Transformation*, 21 YALE J.L. & TECH. (SPECIAL ISSUE) 122 (2019) (identifying four ways AI may transform medical care); Hanchen Wang, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak, Shengchao Liu, Peter Van Katwyk, Andreea Deac, Anima Anandkumar, Karianne Bergen, Carla P. Gomes, Shirley Ho, Pushmeet Kohli, Joan Lasenby, Jure Leskovec, Tie-Yan Liu, Arjun Manrai, Debora Marks, Bharath Ramsundar, Le Song, Jimeng Sun, Jian Tang, Petar Veličković, Max Welling, Linfeng Zhang, Connor W. Coley, Yoshua Bengio & Marinka Zitnik, *Scientific Discovery in the Age of Artificial Intelligence*, 620 NATURE 47 (2023) (examining the various ways AI is increasingly being used to augment and accelerate scientific research); Francesco Filippucci, Peter Gal, Cecilia Jona-Lasinio, Alvaro Leandro & Giuseppe Nicoletti, *The Impact of Artificial Intelligence on Productivity, Distribution and Growth: Key Mechanisms, Initial Evidence and Policy Challenges* 15–37 (OECD Artificial Intelligence Papers No. 015, 2024), https://www.oecd.org/en/publications/the-impact-of-artificial-intelligence-on-productivity-distribution-and-growth_8d900037-en.html (highlighting early evidence of AI's potential for short-term, firm-level productivity gains while acknowledging that long-term macro-level gains will depend on numerous conditions being realized); NAT'L SCI. & TECH. COUNCIL & U.S. DEP'T OF TRANSP., ENSURING AMERICAN LEADERSHIP IN AUTOMATED VEHICLE TECHNOLOGIES: AUTOMATED VEHICLES 4.0, at 2–3 (2020)

a host of new individual and societal risks.[2] In this high-stakes environment, policymakers face difficult institutional choices about how to structure AI governance so as to enable the benefits of continued AI innovation and deployment while also mitigating the risks of harm.

Standards offer a potential answer to this governance question. From the internet to mobile wireless networks to cybersecurity and encryption, standards have emerged over the past few decades as the dominant modality for governing digital technologies.[3] These mutually agreed-upon specifications defining how a technology should function, perform, and be designed have become key instruments for building order within their respective technological domains. Consistent with a body of literature underscoring how design choices can regulate digital technologies, standards can embed constraints directly into a technology's architecture, shaping how it is used and operates in the world.[4] Not all aspects of standards are technical, however. They can also be directed at organizational practices surrounding a technology's development and use, shaping how firms interact with and oversee the systems they deploy.

Building upon their established role in governing prior digital technologies, standards are increasingly being explored as mechanisms for governing AI.[5] Together, AI standards of both the technical and nontechnical variety have the capacity to shape and constrain how AI systems are developed, tested, deployed, and managed over their lifecycle, steering them towards socially desirable ends. Perhaps more importantly, they can do so in a manner that

(highlighting a variety of safety, economic, and social benefits presented by AI-powered autonomous vehicles); Carlos Mureithi, *High Tech, High Yields? The Kenyan Farmers Deploying AI to Increase Productivity*, GUARDIAN (Sep. 30, 2024), https://www.theguardian.com/world/2024/sep/30/high-tech-high-yields-the-kenyan-farmers-deploying-ai-to-increase-productivity (discussing how small-scale farmers in parts of sub-Saharan Africa are utilizing AI tools to enhance crop yields).

2. *See generally* MIA HOFFMANN & HEATHER FRASE, ADDING STRUCTURE TO AI HARM: AN INTRODUCTION TO CSET'S AI HARM FRAMEWORK 12–13 (2023), https://cset.georgetown.edu/publication/adding-structure-to-ai-harm/ (establishing a taxonomy of potential tangible and intangible AI-related harms).

3. For a concise yet helpful overview of standards and their role in governing digital technologies, see *Digital Standards*, GENEVA INTERNET PLATFORM DIGITAL WATCH OBSERVATORY, https://wp.dig.watch/topics/digital-standards (last visited July 8, 2025).

4. *See generally, e.g.*, Lawrence Lessig, CODE AND OTHER LAWS OF CYBERSPACE 6 (1999); Joel R. Reidenberg, *Lex Informatica: The Formulation of Information Policy Rules Through Technology*, 76 TEX. L. REV. 553 (1998); Daniel Benoliel, *Technological Standards, Inc.: Rethinking Cyberspace Regulatory Epistemology*, 92 CALIF. L. REV. 1069 (2004); Deirdre K. Mulligan & Kenneth A. Bamberger, *Saving Governance-by-Design*, 106 CALIF. L. REV. 697 (2018).

5. *See infra* Part II (identifying several ongoing efforts to develop AI standards across various fora).

offers several key advantages.[6] As this Article explores in greater detail below, the fact that standards are typically developed through collaborative, multistakeholder processes makes them well-positioned to draw directly on the expertise of those closest to the technology.[7] This proximity to the developments on the ground also allows for rapid responses to changes in the technological landscape.[8] Moreover, the voluntary, market-driven nature of standards adoption can foster parallel experimentation and competition among prospective standards, enabling the ecosystem to learn from and choose among the most effective practices.[9] Standards can also easily scale across borders, serving as a form of transnational coordination in an increasingly global AI landscape.[10] For these reasons and others, this Article holds that a private standards-based regime currently represents a highly promising approach to AI governance.

To be clear, the argument here is not that standards represent the *perfect* or *optimal* solution to the AI governance question. In the case of AI, a hypothetical "first-best" solution would be democratically accountable, expertly informed, and highly effective, striking a perfect balance between upholding public values and mitigating AI risks on one hand, and fostering innovation and permitting beneficial AI uses on the other. It would articulate expectations for responsible AI development and deployment in terms specific enough to provide clear guidance to AI actors and appropriately tailored to the unique contextual demands of different types of AI systems or use cases. It would be capable of responding swiftly to changing conditions as AI technology continues to evolve, doing so without the need to resort to advance speculation. Finally, it would be able to meaningfully monitor and enforce compliance at relatively low costs and without demanding significant expansions in capacity.

Although standards largely perform well in these dimensions, they do not do so perfectly. Then again, no institutional arrangement could. The reality is that many of the desirable governance attributes above are in tension with one another. For example, the mechanisms for ensuring democratic accountability often come at the expense of speed and responsiveness. Similarly, technically

6. Many of these advantageous features have been discussed at length by scholars examining the "New Governance" model, an alternative regulatory paradigm under which private standards-setting can be located. *See generally, e.g.*, Orly Lobel, *The Renew Deal: The Fall of Regulation and the Rise of Governance in Contemporary Legal Thought*, 89 MINN. L. REV. 342 (2004); Kenneth W. Abbott & Duncan Snidal, *Strengthening International Regulation Through Transnational New Governance: Overcoming the Orchestration Deficit*, 42 VAND. J. TRANSNAT'L L. 501 (2009).

7. *See infra* Section III.B.

8. *See infra* Section III.C.

9. *See infra* Section III.A.

10. *See infra* Section III.D.

detailed and context-sensitive specifications are valuable insofar as they provide clear, actionable guidance to AI developers and users, yet they are also more prone to obsolescence in the face of rapid technological advancements and are more difficult to centrally maintain. Those with the deepest technical expertise frequently have interests or incentives that diverge from the broader public good, while those best equipped to represent societal values tend to lack the technical fluency needed to develop concrete, effective governance measures. Economist Harold Demsetz's pathbreaking work on the "nirvana" fallacy cautions against dismissing a particular governance arrangement because it falls short of some unattainable ideal.[11] Standards should not be rejected as a governance modality simply because they fail to measure up against a benchmark that, while normatively attractive, is not feasible in practice.

Instead, the more appropriate inquiry adopts a comparative second-best approach that compares private standards-based governance to institutional alternatives as they exist in the real world. A complete comparative second-best analysis would consider the full combinatorial suite of possible institutional arrangements and configurations. However, given the abbreviated format of this Article, it mostly narrows its focus to comparing private standards-based governance with the alternative that seems to be at the forefront of the AI governance debate: traditional government regulation. Such regulation is exemplified by several newly adopted state-level AI statutes[12] and by the European Union's Artificial Intelligence Act, the latter of which is being held up as a model for other countries to follow.[13] Yet, even within this limited frame, the relative advantages of private standards-based governance come into clear view. This does not mean, of course, that private standards-based governance is without challenges or tradeoffs; those certainly exist, whether it be the nonbinding nature of standards, potential democratic deficit in their development, or susceptibility to industry capture. However, there are concrete actions that can be taken—drawing lessons from the governance of other digital technologies on how to design legitimate, effective

11. Harold Demsetz, *Information and Efficiency: Another Viewpoint*, 12 J.L. & ECON. 1, 1–4 (1969).

12. *See generally, e.g.*, Act Concerning Consumer Protections in Interactions with Artificial Intelligence Systems, ch. 198, 2024 Colo. Sess. Laws 1199 (codified at COLO. REV. STAT. § 6-1-1701); Artificial Intelligence Policy Act, ch. 186, 2024 Utah Laws (codified at UTAH CODE §§ 13-2-12, 13-70-101 to -305, 76-2-107 and as amended at §§ 13-11-4, 13-61-101, 63I-2-213).

13. *See generally* Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), 2024 O.J. (L 1689) 1.

multistakeholder standards development processes—to make these tradeoffs more favorable such that the benefits of private standards-based governance far outweigh the costs.[14]

The remainder of this Article proceeds as follows. Part II provides a brief overview of AI standards, their functions, and current standardization efforts underway. Part III discusses the comparative strengths of private standards-based governance across four critical dimensions—governance architecture, stakeholder involvement, agility, and scale—while also considering the tradeoffs involved and how they might be better managed. Part IV concludes by offering some final thoughts about the best way forward for AI governance.

II. AI STANDARDS: A BRIEF OVERVIEW

Standards, in the most basic terms, are commonly agreed-upon ways of doing something, whether that be designing a product, organizing information, or performing an activity. Standards serve the critical functions of facilitating interoperability, compatibility, consistency, and quality across different products, systems, and organizations. By defining criteria that enable the formation of shared expectations, standards help reduce uncertainty and transaction costs between parties as well as facilitate consumer trust. They often take the form of *technical standards*: documents specifying common interface designs enabling components to interconnect,[15] formats for structuring information to enable shared understandings,[16] rules for how a technology should behave or interact under different scenarios,[17] or uniform benchmarks and procedures for testing reliability and safety.[18] They can also

14. The term "multistakeholder" lacks a universally accepted definition, but it generally entails two or more classes of actors with shared authority in a common governance enterprise addressing an issue of traditional public concern. *See* Mark Raymond & Laura DeNardis, *Multistakeholderism: Anatomy of an Inchoate Global Institution*, 7 INT'L THEORY 572, 573–82 (2015).

15. *See, e.g.*, *Universal Serial Bus Type-C Cable and Connector Specification Release 2.2*, USB IMPLEMENTERS F., INC. (Oct. 18, 2022), https://usb.org/document-library/usb-type-cr-cable-and-connector-specification-release-22 [https://web.archive.org/web/20221120045921/https://usb.org/document-library/usb-type-cr-cable-and-connector-specification-release-22].

16. *See, e.g.*, *HTML Living Standard*, WEB HYPERTEXT APPLICATION TECH. WORKING GRP., https://html.spec.whatwg.org/multipage/ (last visited Apr. 16, 2025); *Extensible Markup Language (XML) 1.0 (5th ed.)*, WORLD WIDE WEB CONSORTIUM (Nov. 26, 2008), https://www.w3.org/TR/xml/.

17. *See generally, e.g.*, Mike Belshe, Roberto Peon & Martin Thomson, *Hypertext Transfer Protocol Version 2 (HTTP/2)*, INTERNET ENG'G TASK FORCE (May 2015), https://datatracker.ietf.org/doc/html/rfc7540.

18. *See generally, e.g.*, INST. OF ELEC. & ELEC. ENG'RS, *IEEE 3130-2024: Standard for Security Requirements and Testing Methods of Operating Systems in Connected Vehicles* (2024), https://standards.ieee.org/ieee/3130/h10757/.

take the form of *governance* or *management standards*, which are standards that specify organizational structures, processes, and practices for managing internal assets, capabilities, and risk as well as for meeting legal or ethical obligations.[19] Standards, whether they are of the technical or management variety, also vary in their intended scope of applicability: some are general and broadly applicable across domains, whereas others are more domain-specific.

Standards typically emerge through deliberative, consensus-based processes at institutions known as Standards Development Organizations (SDOs). To be sure, a particular technology or product design created unilaterally by a single firm can achieve the status of *de facto* standard simply by becoming widely accepted in the marketplace. An example relevant to the AI context would be NVIDIA's CUDA, a GPU-accelerated computing platform that, at one point, was virtually indispensable (i.e., "the industry standard") for advanced model training and inference.[20] Still, most standards are developed cooperatively in an institutional setting, though the institutions themselves can differ in many important respects. SDOs vary in terms of their composition, with some taking the form of industry consortia made up of mostly corporate stakeholders, while many others are multistakeholder bodies that engage civil society, academia, and technical community in addition to industry. SDOs also vary in terms of geographic scope, ranging from nationally accredited bodies with geographically limited activities to more globally focused bodies.[21] Furthermore, each has its own unique rules and processes governing participation, agenda-setting, voting, and intellectual property rights, all of which can significantly influence the content and adoption of resulting

19. *See generally, e.g.*, INT'L ORG. FOR STANDARDIZATION, *ISO/IEC 27001: Information Security, Cybersecurity and Privacy Protection—Information Security Management Systems—Requirements* (2022), https://www.iso.org/standard/82875.html.

20. *See* Hasan Chowdhury, *CUDA Is Nvidia's Secret Sauce—and Now It's in the Sights of European Regulators*, BUS. INSIDER (July 2, 2024), https://www.businessinsider.com/nvidia-secret-sauce-regulators-gpu-chips-jensen-huang-2024-7. *But see* Anton Shilov, *DeepSeek's AI Breakthrough Bypasses Industry-Standard CUDA for Some Functions, Uses Nvidia's Assembly-Like PTX Programming Instead*, TOM'S HARDWARE (Jan. 28, 2025), https://www.tomshardware.com/tech-industry/artificial-intelligence/deepseeks-ai-breakthrough-bypasses-industry-standard-cuda-uses-assembly-like-ptx-programming-instead (reporting a breakthrough from Chinese firm DeepSeek that has cast doubt on CUDA's continued indispensability).

21. It is not uncommon for an SDO to have either a loosely defined geographic scope or none at all. For example, the Internet Engineering Task Force (IETF), the organization that has historically been responsible for standardizing core and application-layer internet protocols, has neither a physical headquarters nor a formal membership requirement; standards work occurs mainly through electronic mailing list discussions or tri-annual in-person meetings that rotate between host cities, both of which are open to anyone who wishes to participate. Its "geographic scope" is thus a function of where its participants happen to be located, something that has become increasingly diverse over time.

standards.[22] However, what all of these SDOs have in common is that, unlike traditional laws and regulations, the standards they produce are voluntary, meaning adoption is primarily driven by market forces rather than state mandates.

Before proceeding any further, there are a few points worth clarifying. First, standards are distinct from the numerous AI ethical frameworks and high-level principles that have sprung up over the past decade. Much of the supra- and non-national cooperation around AI governance to date has indeed consisted of articulating broad principles—accountability, transparency, fairness, robustness, etc.—that should guide the development and use of AI.[23] While these efforts have been well-intentioned and constitute a good first step towards laying a shared normative foundation, abstract principles alone say little about how AI systems should be developed or used in practice, limiting their effectiveness as AI governance mechanisms.[24] Other commentators have struck a more critical tone, accusing these principles of enabling a form of ethics washing: a rhetorical tool for convincing regulators and the public that concerns are being addressed, yet vague and unenforceable enough so as to require industry to make few meaningful commitments or changes.[25] Regardless of one's perspective on these previous efforts, it is important to recognize that a standards-based regime is not just a continuation of the

22. For a comprehensive comparison of the various rules governing major technical SDOs, see generally Justus Baron, Jorge Contreras, Martin Husovec, Pierre Larouche & Nikolaus Thumm, *Making the Rules: The Governance of Standard Development Organizations and Their Policies on Intellectual Property Rights*, JOINT RSCH. CTR. SCI., EUR. COMM'N EUR 29655 EN (Nikolaus Thumm ed., 2019).

23. *See, e.g.*, Organisation for Economic Co-operation and Development [OECD], *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449 (adopted May 22, 2019; amended May 3, 2024), https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449; United Nations Educational, Scientific and Cultural Organization [UNESCO], *Recommendation on the Ethics of Artificial Intelligence* (adopted Nov. 23, 2021), https://unesdoc.unesco.org/ark:/ 48223/pf0000381137; G20, *G20 Ministerial Statement on Trade and Digital Economy*, annex (June 9, 2019), https://www.mofa.go.jp/policy/economy/g20_summit/osaka19/pdf/documents/en/annex_08.pdf.

24. *See* Brent Mittelstadt, *Principles Alone Cannot Guarantee Ethical AI*, 1 NATURE MACH. INTEL. 501, 503–04 (2019); Jess Whittlestone, Rune Nyrup, Anna Alexandrova & Stephen Cave, *The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions*, 2019 AAAI/ACM CONF. ON AI, ETHICS, & SOC'Y 195, 196–97 (2019), https://doi.org/10.1145/3306618.3314289.

25. *See, e.g.*, Ryan Calo, *Artificial Intelligence and the Carousel of Soft Law*, 2 IEEE TRANSACTIONS ON TECH. & SOC'Y 171, 172 (2021), https://doi.org/10.1109/TTS.2021.3136025; Luke Munn, *The Uselessness of AI Ethics*, 3 AI & ETHICS 869, 872 (2023), https://doi.org/10.1007/s43681-022-00209-w; Karen Hao, *In 2020, Let's Stop AI Ethics-Washing and Actually Do Something*, MIT TECH. REV. (Dec. 27, 2019), https://www.technologyreview.com/2019/12/27/57/ai-ethics-washing-time-to-act/.

"abstract principle" approach to governance. Rather, the point of AI standards is to operationalize these high-level principles, translating them into concrete, technically implementable measures that can be objectively assessed and verified.[26]

This understanding of standards as concrete and well-specified may be counterintuitive to many lawyers, as it represents the inverse of the way the term "standard" has long been employed within the legal tradition.[27] Here, standards have historically been contrasted with rules—the former being open-ended (e.g., reasonableness in tort law, materiality in securities law, or good faith in contract law) and applied ex post and the latter being highly precise and specific and determined ex ante, leaving much less discretion to the party applying the rule.[28] Despite this terminological confusion, the type of technical and governance standards discussed in this Article are, in terms of substance, much more akin to legal rules: they aim to provide specific, actionable ex ante guidance rather than broad discretionary principles.

With that understanding in mind, there are several places where AI standardization activities could beneficially focus.[29] One is the development of different procedures and benchmarks for evaluating model robustness, security, and/or bias—both before deployment and throughout the model's lifecycle as conditions change.[30] These types of standards will almost surely be

26. Some may view this project with skepticism, arguing that framing complex ethical challenges as matters of technical design and implementation reinforces a kind of "technological solutionism" that obscures and sidesteps deeper normative questions. *See, e.g.*, Mittelstadt, *supra* note 24, at 505. Though these concerns are not without merit, nothing about translating principles into implementable practices requires ignoring or oversimplifying the normative questions involved, treating them as purely technical matters that should be addressed exclusively by those with relevant technical expertise. To the contrary, well-designed standards processes can provide structured and inclusive forums for stakeholders to negotiate how values like fairness, accountability, or transparency should be understood and expressed in different technical and organizational contexts.

27. Cass R. Sunstein, *Problems with Rules*, 83 CALIF. L. REV. 953, 959 (1995) ("Lawyers have customarily compared standards ('do not drive unreasonably fast') to rules ('do not go over 60 miles per hour'), with rules seeming hard and fast, and standards seeming open-ended.").

28. *See id.*; Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 DUKE L.J. 557, 559–60 (1992).

29. Many of the areas highlighted here overlap with those outlined in a recent NIST report on the future of U.S. standards engagement. *See generally* JESSE DUNIETZ, ELHAM TABASSI, MARK LATONERO & KAMIE ROBERTS, NAT'L INST. OF STANDARDS & TECH., NIST TRUSTWORTHY AND RESPONSIBLE AI 100-5: A PLAN FOR GLOBAL ENGAGEMENT ON AI STANDARDS (2024), https://www.nist.gov/publications/plan-global-engagement-ai-standards [https://doi.org/10.6028/NIST.AI.100-5] [hereinafter NIST AI Standards Plan].

30. *See id.* at 10–11; *see also* NAT'L TELECOMMS. & INFO. ADMIN., U.S. DEP'T OF COM., NTIA ARTIFICIAL INTELLIGENCE ACCOUNTABILITY POLICY REPORT 26–27 (2024), https://

highly application- or domain-specific. Automated driving systems, for instance, demand much different testing and validation protocols than AI-based medical diagnostic systems.

Another place where standardization efforts are likely to be directed is the definition of standardized formats for disclosing important information about a model, including that about the underlying architecture, training data used, benchmarking/testing results, and any known limitations.[31] Such disclosures help enable downstream actors to accurately assess the model's suitability for specific applications and implement complementary safeguards to enhance overall system safety and reliability.

Similarly, a standardized protocol for reporting discovered model flaws back to the original developer and any affected downstream actors would help facilitate the remediation of potential ecosystem-wide risks in a responsible and streamlined manner.[32] There are also places where AI standards may not be imminent but may emerge in the future as the technology matures. For example, as advancements continue to be made in areas such as model interpretability and explainability, related techniques and technical mechanisms could become promising candidates for standardization activities.[33]

Just as technical standards can help ensure AI systems function reliably and safely, management and governance standards can help address the human and organizational aspects of AI development, use, and oversight. These standards might include—among many other things—implementation best

---

www.ntia.gov/sites/default/files/ntia-ai-report-final.pdf (identifying appropriate information flows to downstream actors as "a critical input to AI accountability").

31. Model cards, including those published by organizations like OpenAI, offer a promising starting point for model disclosures. *See, e.g.*, OpenAI, *GPT-4o System Card* (Aug. 8, 2024), https://openai.com/index/gpt-4o-system-card/. However, such disclosures are not standardized across organizations, vary widely in scope and granularity, and often lack sufficient detail to enable meaningful third-party evaluation or implementation of downstream safeguards.

32. *See* Shayne Longpre, Kevin Klyman, Ruth E. Appel, Sayash Kapoor, Rishi Bommasani, Michelle Sahar, Sean McGregor, Avijit Ghosh, Borhane Blili-Hamelin, Nathan Butters, Alondra Nelson, Amit Elazari, Andrew Sellars, Casey John Ellis, Dane Sherrets, Dawn Song, Harley Geiger, Ilona Cohen, Lauren McIlvenny, Madhulika Srikumar, Mark M. Jaycox, Markus Anderljung, Nadine Farid Johnson, Nicholas Carlini, Nicolas Miailhe, Nik Marda, Peter Henderson, Rebecca S. Portnoff, Rebecca Weiss, Victoria Westerhoff, Yacine Jernite, Rumman Chowdhury, Percy Liang & Arvind Narayanan, *In-House Evaluation Is Not Enough: Towards Robust Third-Party Flaw Disclosure for General-Purpose AI* (Mar. 25, 2025), https://arxiv.org/abs/2503.16861 (identifying the need for a standardized AI flaw report template and coordinated disclosure process).

33. *See* NIST AI Standards Plan, *supra* note 29, at 13 (identifying explainability and interpretability as areas where standards are needed, but that still require "significant foundational work").

practices, internal accountability and governance structures, frameworks in areas such as data quality management or secure development, guidelines for responsible AI procurement, methodologies for conducting impact assessments, and incident response protocols for AI system failures. Sector-specific governance standards may further tailor these approaches to ensure AI deployment aligns with the unique legal, ethical, and operational expectations of different industry verticals.

Many stakeholders have already recognized the need for AI standards, as evidenced by the numerous standardization efforts underway across various organizations and governance levels.[34] A joint technical committee (JTC) of the International Organization for Standardization (ISO) and the International Electrotechnical Committee (IEC), bodies comprised of 165 member countries each represented by a designated national standards organization, launched the earliest coordinated effort to develop global AI standards.[35] Since 2017, ISO/IEC JTC1 has developed and issued dozens of AI standards, though most of its focus thus far has been on the governance and management side as well as on defining foundational concepts and terminology.[36] The Institute of Electrical and Electronics Engineers Standards Association (IEEE-SA), a global multistakeholder body, has also initiated work on a mix of different terminological, technical, and governance standards.[37] At the regional and national levels, U.S. National Institute of Standards and Technology (NIST) has continued to build around its AI Risk Management Framework (RMF) in addition to launching various technical evaluation and testing initiatives.[38] Similarly, the European Committee for Standardization

34. The UK's Alan Turing Institute hosts an extensive and continuously updated online database that compiles both published and in-progress AI standards from various organizations worldwide. *See* ALAN TURING INST., AI STANDARDS HUB, https://aistandardshub.org/ (last visited Apr. 16, 2025).

35. *See* Wael William Diab & Mike Mullane, *How the ISO and IEC Are Developing International Standards for the Responsible Adoption of AI*, UNESCO (Aug. 2, 2024), https://www.unesco.org/en/articles/how-iso-and-iec-are-developing-international-standards-responsible-adoption-ai.

36. *See* NIST AI Standards Plan, *supra* note 29, at 23–27.

37. *See, e.g.*, INST. OF ELEC. & ELEC. ENG'RS, *IEEE 7000-2021: Standard Model Process for Addressing Ethical Concerns During System Design* (2021), https://standards.ieee.org/ieee/7000/6781/; INST. OF ELEC. & ELEC. ENG'RS, *IEEE 3119-2025: Approved Draft Standard for the Procurement of Artificial Intelligence and Automated Decision Systems* (approved Mar. 27, 2025), https://standards.ieee.org/ieee/3119/10729/; INST. OF ELEC. & ELEC. ENG'RS, *IEEE 3129-2023: Standard for Robustness Testing and Evaluation of Artificial Intelligence (AI)-Based Image Recognition Service* (2023), https://standards.ieee.org/ieee/3129/10747/.

38. *See* NAT'L INST. OF STANDARDS & TECH., NIST AI 100-1: ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (2024), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf; *see also, e.g.*, NAT'L INST. OF STANDARDS & TECH., AI RISK

and the European Committee for Electrotechnical Standardization (CEN/CENELEC) have been charged with developing "harmonized standards" to support that implementation of the European Union's landmark AI Act.[39]

Beyond these legacy SDOs, several new fora have emerged to address specific aspects of AI governance. The Partnership on AI is a coalition of technology companies and civil society groups focusing on actionable guidance around the ethical development and use of AI.[40] MLCommons is another multistakeholder consortium that develops performance, quality, and safety benchmarks for AI systems.[41] The Frontier Model Forum is a small alliance of leading frontier AI labs dedicated to identifying best practices for the safe and secure development of large advanced AI models.[42] The Coalition for Content Provenance and Authenticity (C2PA) is an industry-led consortium that creates technical standards for certifying the source of media content amid the rise of deepfakes.[43] All of these initiatives reflect a growing recognition that effective AI governance requires more than just high-level principles, but also concrete steps and measures for putting those principles in practice. As AI capabilities continue to advance and AI systems become more deeply embedded into economic and social life, the role of standards will only grow in importance.

III. PRIVATE STANDARDS-BASED GOVERNANCE: PROMISES AND SURMOUNTABLE CHALLENGES

As AI systems become increasingly pervasive, more granular articulations of expected or desired behavior surrounding AI will become a virtual necessity. There are several forces that may contribute to this need. It could be the demand for common baselines against which ex ante audits or tests can be conducted, a way to foster trust throughout the ecosystem by signaling

MANAGEMENT FRAMEWORK PLAYBOOK (2023), https://airc.nist.gov/docs/AI_RMF_Playbook.pdf.

39. *See* Hadrien Pouget & Ranj Zuhdi, *AI and Product Safety Standards Under the EU AI Act*, CARNEGIE ENDOWMENT FOR INT'L PEACE (Mar. 5, 2024), https://carnegieendowment.org/research/2024/03/ai-and-product-safety-standards-under-the-eu-ai-act?lang=en.

40. *About Us*, P'SHIP ON AI, https://partnershiponai.org/about (last visited Apr. 16, 2025).

41. *About Us: An Open AI Engineering Consortium*, MLCOMMONS, https://mlcommons.org/about-us/ (last visited Apr. 16, 2025).

42. *About Us*, FRONTIER MODEL F., https://www.frontiermodelforum.org/about-us/#mission (last visited Apr. 16, 2025).

43. *Coalition for Content Provenance and Authenticity (C2PA) Charter*, COAL. FOR CONTENT PROVENANCE & AUTHENTICITY, https://c2pa.org/about/charter/ (last visited Apr. 16, 2025).

information about a model's quality or the adequacy of an organization's compliance efforts. Similarly, it could be the demand for clear frameworks that help guide ex post determinations of responsibility when AI-enabled harms occur. Perhaps detailed specifications are sought to facilitate vertical coordination and enable firms at different levels of the supply chain to jointly provide AI-based products or services in a safe and reliable fashion. It could even be as simple as the need to codify knowledge about best practices for AI risk mitigation so that firms have actionable guidance as they navigate an uncertain technological frontier.

Regardless of what the main driver is, standards-like specifications for AI—whether promulgated by a legislative or regulatory body or inferred ex post from a body of decisions assessing liability—are an inevitability. One of the few questions that remains is who should determine their content. The question of "who decides," as legal scholar Neil Komesar explains in his book *Imperfect Alternatives*, is fundamentally a question of institutional choice.[44] The answer to this question may ultimately dictate the degree to which social and policy goals around AI are realized.[45]

Among the institutional alternatives available for governing AI, the most promising—particularly in its ability to balance AI harm prevention with continued innovation and deployment—is private standards-based governance. To reiterate an important point raised in the introduction, this Article does not argue that private standards-based governance represents the *optimal* or *ideal* solution to the AI governance puzzle. It is extremely unlikely that such a "first-best" solution exists. In what follows, this Article examines the relative strengths of standards across the four critical dimensions mentioned briefly above—governance architecture, stakeholder involvement, agility, and scale—while also considering the tradeoffs that exist and how they might be better managed. When compared to other institutional arrangements, each viewed in light of the real-world limitations and the tradeoffs they entail, private standards-based governance emerges as the most desirable of the second-best options.

A. GOVERNANCE ARCHITECTURE

1. The Benefits of Bottom-Up, Decentralized Governance

Classic command-and-control regulation seeks to achieve outcomes by issuing legal commands—typically in the form of prescribed conduct or technical configurations—that leave regulated entities with little discretion

44. NEIL K. KOMESAR, IMPERFECT ALTERNATIVES: CHOOSING INSTITUTIONS IN LAW, ECONOMICS, AND PUBLIC POLICY 3 (1994).

45. *See id.* ch. 2.

over implementation.[46] While this can provide regulated entities with greater certainty from a compliance standpoint, the requirements issued by regulators may be suboptimal or even ineffective.[47] This is particularly cause for concern given governments' less-than-stellar track record in setting or selecting appropriate technical standards in high-technology domains.[48] Flaws in government-mandated AI specifications can be costly, undermining both AI innovation and the effectiveness of governance efforts, as the binding nature of these specifications compels adoption regardless of any shortcomings.[49]

By contrast, the voluntary market-driven nature of private standards adoption allows for more meritocratic selection of the best approaches. When multiple competing standards emerge—and they frequently do—decisions about which standards prevail are made through the decentralized choices of developers, users, and other ecosystem participants.[50] To gain traction in the marketplace in the form of adoption, private standards must prove their value

46. *See* Kenneth A. Bamberger, *Regulation as Delegation: Private Firms, Decision Making, and Accountability in the Administrative State*, 56 DUKE L.J. 377, 386 (2006) [hereinafter Bamberger, *Regulation as Delegation*]; Lobel, *supra* note 6, at 376 ("Their agency is limited to choosing whether to comply with the regulations to which they are subjected.").

47. *See* Daniel Gervais, *The Regulation of Inchoate Technologies*, 47 HOU. L. REV. 665, 704 (2010).

48. For surveys of past efforts showing the shortcomings of government-set standards, see STANLEY M. BESEN & LELAND L. JOHNSON, COMPATIBILITY STANDARDS, COMPETITION, AND INNOVATION IN THE BROADCASTING INDUSTRY 135 (1986); JEFFREY H. ROHLFS, BANDWAGON EFFECTS IN HIGH-TECHNOLOGY INDUSTRIES 201 (2001). One of the most notorious examples was the U.S. government's initiative to standardize a hardware-based key escrow system—better known as the "Clipper Chip"—that would permit law enforcement to obtain "exceptional access" to encrypted communication. The system saw little adoption and was abandoned just a few years later, due in no small part to a widely cited paper revealing significant technical vulnerabilities. *See generally* Matt Blaze, *Protocol Failure in the Escrowed Encryption Standard*, 2 PROC. ASS'N FOR COMPUTING CONF. ON COMPUT. & COMMC'NS SEC. 59 (1994) (finding major security flaws in the government's Escrowed Encryption Standard); *see also* Stacy Baird, *The Government at the Standards Bazaar*, 18 STAN. L. & POL'Y REV. 35, 67–70 (2007) (examining several other instances of "government failure" in technical standards setting); STEPHEN BREYER, REGULATION AND ITS REFORM 115 (1982) (identifying several challenges that traditional regulators face in attempting to set standards).

49. BREYER, *supra* note 48, at 102.

50. The emphasis placed on parallel experimentation and competition between prospective standards has often been described as a uniquely American approach to standardization. *See* DIETER ERNST, EAST-WEST CTR., POLICY STUDIES NO. 66, AMERICA'S VOLUNTARY STANDARDS SYSTEM: A 'BEST PRACTICE' MODEL FOR ASIAN INNOVATION POLICIES? 32 (Edward Aspinall & Dieter Ernst eds., 2013); TIM BÜTHE & WALTER MATTLI, THE NEW GLOBAL RULERS: THE PRIVATIZATION OF REGULATION IN THE WORLD ECONOMY 162 (2011). However, it is also the model that the digital technology standards-setting landscape has come to mostly closely resemble—perhaps a reflection of the United States' outsized influence in this domain.

to prospective implementers.[51] The choices the market ultimately converges around can be difficult to predict in advance. For example, there was a point in time when many expected Bluetooth would become the dominant wireless technology for connecting user devices to local area networks (among its many other uses). Yet, it was the superior range and transfer speeds provided by the IEEE 802.11 standard—known commercially as Wi-Fi—that ultimately caused it to triumph over the low-cost, low-power consumption Bluetooth.[52]

The same competitive dynamic that exists between rival standards can also extend to the SDO level, as standard-setting venues themselves must vie to attract and retain participants in order to stay relevant.[53] This process resembles Charles Tiebout's famous model under which interjurisdictional competition can lead to the efficient provision of local public goods.[54] Just as residents in the Tiebout model are free to move to communities that match their preferred level of local public good provision,[55] stakeholders in the standardization process have the opportunity to "vote with their feet" in choosing where to participate. The resulting competitive pressures can help ensure that standard-setting remains responsive to needs rather than stagnating under some combination of complacency and institutional inertia.[56]

An example of this phenomenon can be found in the efforts to standardize the World Wide Web during the mid-1990s. After growing frustrated with the slow pace and "endless philosophical rat holes" they had encountered at the Internet Engineering Task Force (IETF), web pioneer Tim Berners-Lee and several other early browser developers decided to defect and form the World Wide Web Consortium (W3C).[57] While certainly not without faults of its own, the W3C has proven instrumental in the Web's rapid and successful

51. *See* NIST AI Standards Plan, *supra* note 29, at 5.

52. Joseph Farrell & Timothy Simcoe, *Four Paths to Compatibility*, *in* THE OXFORD HANDBOOK OF DIGITAL ECONOMY 34, 40 (Martin Peitz & Joel Waldfogel eds., 2012) ("For example, Bluetooth (IEEE 802.15) was conceived as a home networking standard, but ceded that market to Wi-Fi (IEEE 802.11) and is now widely used in short-range low-power devices, such as wireless headsets, keyboards, and remote controls.").

53. *See* Baron et al., *supra* note 22, at 64–65; *see also* Errol Meidinger, *Competitive Supragovernmental Regulation: How Could It Be Democratic?*, 8 CHI. J. INT'L L. 513, 531 (2008) (making this same observation about the transnational private regulatory landscape more generally).

54. *See generally* Charles M. Tiebout, *A Pure Theory of Local Expenditures*, 64 J. POL. ECON. 416 (1956).

55. *Id.* at 418.

56. *Cf.* Baron et al., *supra* note 22, at 67 (suggesting that inter-venue competition can serve as a check on a given SDO's ability to impose unfavorable new policies in areas such as IP rights).

57. TIM BERNERS-LEE, WEAVING THE WEB: THE ORIGINAL DESIGN AND ULTIMATE DESTINY OF THE WORLD WIDE WEB BY ITS INVENTOR 62 (1999).

standardization, demonstrating the benefits of responsive governance when faced with competitive pressure.

A private standards-based regime would also allow room for early-stage and continuous experimentation, permitting several different approaches to develop in parallel so the ecosystem can learn from different models before either codifying them into standards or deciding to adopt them.[58] Such experimentation is especially valuable given the heterogeneity among AI systems and use cases, which almost certainly renders a one-size-fits-all approach inappropriate.[59] Instead of a regulator attempting to determine *a priori* if and where sector-specific approaches are needed, this differentiation can be dictated by the needs and challenges experienced by those operating within each sector.[60] Groups of related actors such as medical AI developers or autonomous vehicle manufacturers can determine whether a horizontal approach is appropriate for their intended use-case and, if not, split off and develop their own specialized standards that better address domain-specific concerns.

Top-down specifications imposed by regulators, on the other hand, stand to foreclose the possibility of private experimentation around potential improvements or alternatives.[61] In order to avoid noncompliance, regulated entities would need to continue adhering to the mandated specifications until the regulator updates them, even if the specifications grew outdated or were poorly suited for a new use case.[62] Such concerns are particularly acute in the

58. *See* Daniel E. Walters & Hannah J. Wiseman, *Self-Regulation in Emerging and Innovative Industries*, 62 HOU. L. REV. 543, 564–65 (2025); *see also* Lobel, *supra* note 6, at 380, 382 (explaining that the diversity and experimentation enabled under decentralized new governance approaches is more than just a temporary process for identifying and ultimately converging on the best solution, but it also serves as a "means for continuous change and improvement").

59. Cary Coglianese, *Regulating Machine Learning: The Challenge of Heterogeneity*, COMPETITION POL'Y INT'L: TECHREG CHRON. 8 (Feb. 2023) [hereinafter Coglianese, *Machine Learning*]; *see also* Lobel, *supra* note 6, at 379–80 (arguing that governance approaches emphasizing diversification and pluralized solutions are better suited for complex, dynamic regulatory domains than uniform, top-down models).

60. Walters & Wiseman, *supra* note 58, at 571.

61. *See* BREYER, *supra* note 48, at 105 (explaining that government-mandated design standards diminish firms' "incentive to look for better methods"); Lobel, *supra* note 6, at 393 (explaining that one reason we might prefer softer alternatives to traditional hard law is that the former, unlike the later, allows for deviance and trial and error without the fear of sanctions).

62. *See* BREYER, *supra* note 48, at 115–16 (highlighting the tendency of government mandated design standards to "freeze technology"). *But see* Cary Coglianese, *The Limits of Performance-Based Regulation*, 50 U. MICH. J. L. REFORM 525, 542 (2017) (recognizing there are actions that regulators can take to reduce technology "lock-in" and accommodate new

context of AI, a technology most would agree is still in its infancy.[63] It is extremely likely that any initial regulatory specifications would need significant refinement over time as both understanding of AI and the technology itself continue to evolve, yet the limited feedback channels within a command-and-control regime make adaptation significantly more difficult.[64]

Concededly, command-and-control regulation is not the only item traditional regulators have in their toolkit. Performance-based regulation, which focuses solely on outcomes rather than the means of achieving those outcomes, may appear to offer a more flexible yet nonetheless state-centric alternative.[65] Here, regulators define a specific performance target or goal that must be met, but leave it up to each regulated entity to determine how to achieve it.[66] The main weakness of the performance-based approach, however, is that not every area of regulation is conducive to measurable outcomes and AI is arguably one of them.[67] Unlike domains such as air pollution where regulators can set specific, quantifiable emissions thresholds, AI systems present complex, multidimensional risks that often resist simple performance metrics.[68]

Less prescriptive forms of state-centric regulation can also create significant uncertainty by threatening to penalize firms without simultaneously offering them concrete guidance or a means of easily verifying compliance. With a cutting-edge technology like AI, where knowledge pertaining to the most advanced systems has an almost esoteric quality, many firms—small and medium-sized enterprises (SMEs) in particular—are poorly equipped to navigate this uncertainty on their own.[69] Even though the technical expertise

innovations, such as establishing waiver mechanisms that allow firms to depart from otherwise binding specifications).

63. *See* Gervais, *supra* note 47, at 671–73, 702 (arguing that traditional regulatory interventions aimed at inchoate technologies—those that "are far from completely developed" and for whom the "future is unpredictable"—are more likely to miss their targets).

64. *See id.* at 676 ("[T]here are rarely feedback loops to adjust the regulatory framework if the target is missed.").

65. Coglianese, *The Limits of Performance-Based Regulation*, *supra* note 62, at 526.

66. *Id.*

67. *See* Bamberger, *Regulation as Delegation*, *supra* note 46, at 389 ("Certain public problems . . . lend themselves to neither specific behavioral commands nor measurable outcomes."); *see also, e.g.*, David Thaw, *The Efficacy of Cybersecurity Regulation*, 30 GA. ST. U. L. REV. 287, 301–02 (2014) (discussing how the lack of well-defined outcome metrics in information security complicates the application of performance-based regulatory approaches).

68. *See* Coglianese, *Regulating Machine Learning*, *supra* note 59, at 8 ("[I]n many cases it will be unlikely that regulators can develop sufficiently clear, monitorable performance tests for algorithms themselves.").

69. A similar phenomenon of disproportionate compliance burden on smaller entities due to regulatory complexity and uneven distribution of expertise has been observed with the

possessed by private actors is superior to that of regulators *in the aggregate*, this expertise is unevenly distributed: larger firms can afford to employ in-house AI specialists or engage outside consultants, but SMEs are much less likely to possess the resources to determine best practices on their own. As a result, regulatory uncertainty may disproportionately burden SMEs, chilling AI adoption and deployment due to the unpredictable legal and compliance risks they would face.[70]

To briefly summarize thus far: highly concrete, implementable specifications are desirable for effective AI governance, but state-centric modes of regulation are poorly positioned to provide this concreteness and even risk stunting the technology's development by mandating suboptimal approaches. A private standards-based regime, meanwhile, would allow approaches to emerge more organically from the bottom-up. By enabling decentralized experimentation and choice, such a regime fosters more effective, adaptive, and innovation-permitting governance.

Finally, under a private standards-based governance regime, the role of government would undergo a shift towards areas where it retains comparative institutional advantages: enforcing commitments, addressing market failures, and policing anticompetitive conduct within the standardization process. This might also include holding firms accountable ex post when they make representations to the public or other firms about their adherence to particular standards yet fail to do so.[71] The shift in function here reflects Judge Frank Easterbrook's influential insight from the early days of the internet: in the face of rapid technological change and uncertainty, the most productive role for public law is *not* to anticipate the future and attempt to engineer bespoke solutions but to instead support private ordering.[72] In Easterbrook's words, "If you don't know what is best, let people make their own arrangements."[73]

EU's General Data Protection Regulation (GDPR). *See, e.g.*, Sean Sirur, Jason R.C. Nurse & Helena Webb, *Are We There Yet? Understanding the Challenges Faced in Complying with the General Data Protection Regulation (GDPR)*, arXiv:1808.07338v1, at 5–6, 8 (Aug. 22, 2018), https://arxiv.org/pdf/1808.07338v1.

70. *See id.*

71. The Federal Trade Commission (FTC) has broad authority to police unfair and deceptive acts under Section 5 of the FTC Act and has already used it to bring enforcement actions against firms accused of misrepresenting the capabilities of their AI systems. *See* 15 U.S.C. § 45(a)(1); *see also* Press Release, Fed. Trade Comm'n, FTC Announces Crackdown on Deceptive AI Claims and Schemes (Sep. 25, 2024), https://www.ftc.gov/news-events/news/press-releases/2024/09/ftc-announces-crackdown-deceptive-ai-claims-schemes.

72. *See* Frank H. Easterbrook, *Cyberspace and the Law of the Horse*, 1996 U. CHI. LEGAL F. 207, 207–13.

73. *Id.* at 210.

Notably, the re-definition of the government's role would dramatically reduce administrative costs.[74] This is particularly salient given the large number of actors across different industry verticals who are already beginning to develop and use AI systems for a variety of different purposes, as any form of centralized oversight and enforcement will be both challenging and demand significant state resources.[75] While administrative costs are neither the only nor an overriding consideration, they assume heightened significance in the current U.S. political climate, which has seen a sharp rise in hostility towards the administrative state under the second Trump administration.[76] Repositioning government more as a facilitator and backstop instead of the primary standard-setter offers a more resource-efficient pathway to effective AI governance.

2. Second-Best Considerations: Voluntariness and Possible Races to the Bottom

The room for experimentation and competition that exists under a private standards-based regime does not come without tradeoffs. Unlike traditional state-centric regulation, where standards can be given the force of law when widespread adoption is desirable, private standards are still entirely voluntary. The manner in which certain types of AI technical standards enable interoperability, enhance product compatibility, and facilitate multi-firm provision provides strong incentives for adoption. However, other types of AI standards will impose compliance costs on adopting firms without offering an immediate, tangible benefit. For example, standards aimed at minimizing algorithmic bias and discrimination might not directly contribute to a company's bottom line, at least in the short term. Even when standards of the latter variety are developed through multistakeholder, consensus-based

74. *See* Abbott & Snidal, *supra* note 6, at 525 ("Decentralization thus reduces demands on the state, a significant advantage in an era when many states and agencies face both shrinking resources and growing demands for action.").

75. *See* Coglianese, *Regulating Machine Learning*, *supra* note 59, at 12 ("[I]t is machine learning's heterogeneity that poses regulators' greatest challenge of all. These algorithms' varied forms, multiple uses, and dynamic proper ties make most conventional regulatory strategies obsolete."); *see also* Lobel, *supra* note 6, at 396 (explaining that centralized command-and-control oversight is less preferable under conditions of rapid advancement, heterogeneity, and complexity); Richard B. Stewart, *Administrative Law in the Twenty-First Century*, 78 N.Y.U. L. REV. 437, 446 (2003) (arguing that centralized regulation "suffers from the inherent problems involved in attempting to dictate the conduct of millions of actors in a quickly changing and very complex economy and society throughout a large and diverse nation"); Cass Sunstein, *Administrative Substance*, 40 DUKE L.J. 607, 627 (1991) (identifying the use of "highly bureaucratized 'command-and-control' regulation" to regulate a large number of subjects in an diverse country as a "large source of regulatory failure in the United States").

76. *See* Jody Freeman & Sharon Jacobs, *President Trump's Campaign of 'Structural Deregulation,'* LAWFARE (Feb. 12, 2025), https://www.lawfaremedia.org/article/president-trump-s-campaign-of--structural-deregulation.

processes, their nonbinding nature raises a fundamental question: will firms implement them on their own volition? Some self-regulation skeptics are likely to maintain that, absent an enforcement mechanism, a system of voluntary standards will only lead to more of the same "ethics washing" that existing, high-level AI frameworks have been charged with perpetuating.[77]

A closely related challenge is that competitive pressures are not guaranteed to push standardization outcomes in directions that align with public interests. Given that industry actors—to a much greater extent than other stakeholders—will be the ones facing standards adoption decisions and are thus primary arbiters of a standard's value, their criteria for evaluating the attractiveness of a standard will not necessarily be consistent with broader governance objectives. Similarly, in the context of inter-SDO competition, it is possible that standard-setting venues are selected not because they foster high-quality, effective standards but because they cater to specific stakeholder groups seeking to advance their own priorities. Critics may even contend that this invites a race to the bottom, where firms inevitably gravitate toward venues that allow them to shape less stringent standards with weaker protections for public values and interests.[78]

Though the tradeoffs identified above are undoubtedly real, they are not as unfavorable as they first appear, and there are even steps that can be taken to manage them more efficiently. One of the most powerful forces counteracting the risk of weak or inconsistently implemented standards is the overhanging threat of public regulatory intervention. If private governance efforts are perceived as inadequate or as failing to meaningfully address AI risks, regulators may respond by adopting a more top-down, coercive regulatory posture.[79] This latent threat may provide firms with strong incentives to voluntarily adopt more rigorous standards in order to stave off heavier-handed (and potentially disruptive) government action.[80] In fact, harnessing this dynamic and giving the threat some credibility—continuing to

77. *See supra* note 25 and accompanying text.

78. *See* Walters & Wiseman, *supra* note 58, at 562–64 (acknowledging the possibility of a race-to-the-bottom standards and SDO-based competition, though exploring several contextual factors that may mitigate it).

79. *See* IAN AYRES & JOHN BRAITHWAITE, RESPONSIVE REGULATION: TRANSCENDING THE DEREGULATION DEBATE 38–39 (1992) (maintaining that by signaling its willingness to escalate towards more coercive, interventionist regulatory strategies, the government gives industry actors incentives to "make regulation work at lower levels of interventionism"); Abbott & Snidal, *supra* note 6, at 523 (noting that "the threat of [state] intervention reinforces softer New Governance measures.").

80. John W. Maxwell, Thomas P. Lyon & Steven C. Hackett, *Self-Regulation and Social Welfare: The Political Economy of Corporate Environmentalism*, 43 J.L. & ECON. 583, 612–13 (2000) (finding empirical support for this proposition).

monitor the general landscape and periodically assessing the need for intervention—can be one of the most effective tools the government has for inducing firms to act without resorting to mandates.[81]

It is also important not to lose sight of the broader legal backdrop against which debates over AI regulation are taking place. The threat of new binding regulations is not the only reason firms have for taking AI safety and trustworthiness seriously on their own. The development and use of AI systems—at least in the United States—is already subject to several generally applicable, technology-neutral legal and regulatory frameworks.[82] For example, those who delegate hiring, housing, or loan application decisions to AI-based systems must still comply with employment, housing, and lending discrimination laws, respectively.[83] Likewise, when an AI system that is developed or used in a careless manner goes on to cause someone injury, the responsible firm can still be held liable in tort under negligence and/or products liability theories.[84] While existing legal frameworks may not offer a

81. The government acting in this capacity straddles the line between a pure self-regulatory approach and what has been described as "meta-regulation" (i.e., active state oversight and influence over self-regulatory efforts). *See* Cary Coglianese & Evan Mendelson, *Meta-Regulation and Self-Regulation*, *in* THE OXFORD HANDBOOK OF REGULATION 147, 161–62 (Robert Baldwin, Martin Cave & Martin Lodge eds., 2010). This logic also parallels what Tim Wu has identified as the strategic use of "agency threats" (e.g., warning letters, public speeches hinting at possible action, etc.) to influence private behavior absent formal rulemaking, which he argues is most justified when industries are experiencing rapid change and thus a great amount of uncertainty. *See generally* Tim Wu, *Agency Threats*, 60 DUKE L.J. 1841 (2011).

82. *See* Mariano-Florentino Cuéllar, *A Common Law for the Age of Artificial Intelligence: Incremental Adjudication, Institutions, and Relational Non-Arbitrariness*, 119 COLUM. L. REV. 1773, 1781 (2019) ("[S]ociety already 'regulates' AI . . . even in the absence of statutes and regulatory rules governing AI . . . . [T]he ultimate regulatory backstop here is the common law.").

83. *See, e.g.*, Exec. Order No. 14,110, § 7.3, 88 Fed. Reg. 75191, 72213 (Nov. 1, 2023), https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence (instructing the Consumer Financial Protection Bureau and the Department of Housing and Urban Development to provide guidance to the private sector on how to ensure AI-based decision making tools did not violate existing credit and housing discrimination laws); CONSUMER FIN. PROT. BUREAU, U.S. DEP'T OF JUST., EQUAL EMP. OPPORTUNITY COMM'N & FED. TRADE COMM'N, *Joint Statement on Enforcement Efforts Against Discrimination and Bias in Automated Systems* (Apr. 25, 2023), https://www.ftc.gov/system/files/ftc_gov/pdf/EEOC-CRT-FTC-CFPB-AI-Joint-Statement%28final%29.pdf ("[E]xisting legal authorities apply to the use of automated systems and innovative new technologies just as they apply to other practices.").

84. *See* Bryan H. Choi, *Negligence Liability for AI Developers*, LAWFARE (Sep. 26, 2024), https://www.lawfaremedia.org/article/negligence-liability-for-ai-developers (exploring how negligence liability might apply to AI developers).

perfect fit for AI and the process of adaptation is both slow and ongoing,[85] they do present some degree of legal risk for firms. AI standards not only help firms navigate this risk by defining implementable measures for addressing the underlying source (e.g., model flaws, bias), but they can even come to shape the law itself by serving as benchmarks that courts or regulators reference when assessing legal responsibilities. For instance, a firm's adherence to widely accepted, relevant technical standards may serve as evidence (albeit non-dispositive evidence) of reasonableness when facing negligence or defective design claims.[86]

Finally, though true that it is mostly commercial actors that will be developing and deploying AI systems and thus ultimately deciding which standards get adopted—giving them more leverage than other stakeholder groups when "voting with their feet"—this does not mean a race to the bottom is either the inevitable or likely result of inter-SDO competition. Unlike the Tiebout model, where it is assumed that community residents have perfect mobility,[87] there are still some constraints on the ability of industry actors to move freely to new venues that match their individual preferences. Consider a scenario in which industry stakeholders shifted to a more lenient multistakeholder venue that allowed them to develop more relaxed AI standards. Meanwhile, stakeholders from civil society, academia, and the technical community stayed behind and, through a consensus-based deliberative process, developed far more stringent standards designed to better uphold public interests.[88] Those adopting the industry-developed standard would risk facing immense public backlash for consciously spurning a more democratically legitimate alternative in favor of weaker public interest protections—backlash that, in addition to hurting a company's reputation, could culminate in the type of public regulatory interventions firms seek to avoid.[89] This gives commercial actors yet another incentive to cooperate with

85. *See generally, e.g.*, Andrew D. Selbst, *Negligence and AI's Human Users*, 100 B.U. L. REV. 1315 (2020) (identifying several challenges that negligence law poses for plaintiffs seeking redress for AI-related harms).

86. For an overview of how private technical standards interact with the tort system, see generally GARY E. MARCHANT, SWORDS AND SHIELDS: IMPACT OF PRIVATE STANDARDS IN TECHNOLOGY-BASED LIABILITY (2022), https://ssrn.com/abstract=4178750.

87. *See* Tiebout, *supra* note 54, at 419.

88. *Cf.* Gregory C. Shaffer & Mark A. Pollack, *Hard vs. Soft Law: Alternatives, Complements, and Antagonists in International Governance*, 94 MINN. L. REV. 706, 795–98 (2010) (explaining that non-state actors will often attempt to counter transnational soft law instruments they find unfavorable by either developing their own rival soft-law instruments or pushing international organizations, such as the UN, to do so).

89. *See* Julia Black, *Constructing and Contesting Legitimacy and Accountability in Polycentric Regulatory Regimes*, 2 REGUL. & GOVERNANCE 137, 146 (2008) (describing the concept of pragmatic legitimacy, wherein industry actors are motivated solely by economic interests to

other affected stakeholders at more inclusive SDOs: not just to shape the content of emerging standards, but to ensure those standards carry the legitimacy needed to secure public trust and forestall heavier-handed government oversight.

B. STAKEHOLDER INVOLVEMENT

1. Leveraging Field-Level Expertise and Inclusive Participation

Effective AI governance requires a deep understanding of the technology itself. Yet, because AI represents a highly complex and cutting-edge field, regulatory agencies—even specialized ones—inevitably face limitations in their technical capacity and knowledge.[90] That is not to say that governments are incapable of expanding this capacity. Indeed, both the first Trump and Biden administrations prioritized efforts to bring in top AI talent to federal agencies, commencing various initiatives to streamline the hiring process for those with AI-related technical experts.[91] However, competing with the private sector over a highly in-demand talent pool may be unrealistic, as private sector salaries for new AI PhDs can reach as high as $800,000.[92] Further compounding the difficulties here is that AI expertise is often highly domain-specific: different deployment contexts and use cases present unique considerations, meaning that hiring a handful of generalist experts may not be sufficient to address the full range of governance challenges AI presents.[93]

---

comply with private regulatory standards that consumers perceive as legitimate, even when they themselves find the standards normatively undesirable); Meidinger, *supra* note 53, at 525 ("Easy exit, however, is often seriously constrained by practical power. Many firms choose to subject themselves to supragovernmental regulatory standards not so much because they wish to live under them as because they feel that they must in order to avoid significant economic losses.").

90. *See* Gary E. Marchant & Carlos Ignacio Gutierrez, *Soft Law 2.0: An Agile and Effective Governance Approach for Artificial Intelligence*, 24 MINN. J.L. SCI. & TECH. 375, 384 (2023); *see also* Coglianese, *Regulating Machine Learning*, *supra* note 59, at 6 (emphasizing the challenges posed by the domain-specific nature of AI-related technical knowledge, which makes it unlikely that generalist AI agency could ever possess the expertise needed to regulate all of AI's heterogenous uses).

91. *See* Exec. Order No. 13,960, Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government, § 7, 85 Fed. Reg. 78939, 78942 (Dec. 3, 2020); Exec. Order No. 14,110, *supra* note 83, § 10.2.

92. *See* Sam Shedden, *AI Researcher Salary: Eye-Watering Pay of Top Tech Job Revealed*, READWRITE (Jan. 2, 2024), https://readwrite.com/ai-researcher-salary-eye-watering-pay-of-top-tech-job-revealed/; Chris Stokel-Walker, *Regulators Need AI Expertise. They Can't Afford It*, WIRED (Mar. 14, 2024), https://www.wired.com/story/regulators-need-ai-expertise-cant-afford-it/.

93. *See* Coglianese, *Regulating Machine Learning*, *supra* note 59, at 6.

It would not be a stretch to suggest that the technical knowledge possessed by government agencies is unlikely to ever surpass or even match that which is distributed across the private sector.[94] Hence, one of the key advantages of private standards-based governance is its ability to *directly* leverage the collective technical expertise and experience of those working at the forefront of AI development.[95] In addition to the advantages the private sector has in attracting human capital, those actively developing and deploying AI systems possess firsthand experience identifying different risks and failure modes as well as designing mitigation strategies.[96] These actors are also well-positioned to conduct ongoing real-world testing and provide continuous feedback on what works in practice versus what merely sounds good in theory.[97] This continuous cycle of iteration and learning gives private actors a unique advantage in shaping standards that are both technically sound and practically implementable.

When standards development is conducted in an open and cooperative setting, inclusive participation in the process has the potential to yield additional benefits. Allowing regulated entities to meaningfully contribute to the creation of standards can increase industry buy-in and enhance the likelihood that implementation will go beyond mere box-checking.[98] Equally important is the inclusion of non-industry voices in the process—especially civil society stakeholders representing public interests—which tends to strengthen the legitimacy of resulting standards.[99] Legitimacy here operates on two levels: there is what we might call *internal legitimacy*, which is concerned with the perception of participants at a particular venue or prospective adopters, and *external legitimacy*, which is concerned more broadly with whether

94. The EU AI Act's reliance on technical standards implicitly acknowledges the comparative advantage that specialized, multistakeholder bodies possess when it comes to translating broad regulatory goals into concrete technical practices. *See supra* note 39 and accompanying text.

95. *See* Lobel, *supra* note 6, at 382 (explaining that decentralized new governance approaches affirm the principle of subsidiarity: the idea that decisions are best made by those who possess superior information about a problem by virtue of their proximity to it); Abbott & Snidal, *supra* note 6, at 528–29 (noting the new governance approaches recognize "that expertise is often dispersed, and [seek] to harness a wide range of stakeholders who may have 'local' expertise otherwise unavailable to the state.").

96. *See* Walters & Wiseman, *supra* note 58, at 567 (noting that firms are particularly likely to possess superior knowledge about risks in a given industry during the early stages of its development).

97. *See* Meidinger, *supra* note 3, at 529.

98. *See* AYRES & BRAITHWAITE, *supra* note 79, at 113 ("If business is responsible for writing and enforcing its own code of conduct, the notion of regulation may become more palatable.").

99. *See infra* note 112 and accompanying text.

society at large perceives these governance arrangements as acceptable.[100] With AI standards, the latter type is of equal importance, if not even greater.

Why does the general public's perception of AI standards matter? As tempting as it may be to view AI standards as purely technical, this is not necessarily the case: the design of a technical standard can embed normative choices that have social, economic, and political consequences once implemented in the real world.[101] For example, benchmarks for evaluating model bias call for implicit value judgments about acceptable levels of discrimination risk, while transparency standards for disclosing information about a model's architecture or training data must determine how to weigh the benefits of information-sharing against the costs and risks that disclosure poses to the model developer.[102] Delegating highly consequential, value-laden choices to a narrow group of actors without any accountability mechanisms or stakeholder input is unlikely to be a politically tenable arrangement in the long term. Ensuring that diverse perspectives and interests are given meaningful weight in the deliberation process, however, helps mitigate these concerns and fosters legitimacy, making private standards-based governance a more sustainable approach.

2. Second-Best Considerations: Industry Capture and Participation Barriers

While private standards development can better leverage industry expertise in pursuit of more technically informed and effective governance mechanisms, the knowledge asymmetries that exist in such arrangements can be a double-edged sword. Since industry participants—or at least a particular subset of them—tend to possess a much more sophisticated understanding of both the underlying technology and the practical consequences of certain technical design choices, they might use their informational advantages in anticompetitive or self-serving ways.

---

100. *Cf.* Black, *supra* note 89, at 147 (explaining that a regulator's legitimacy depends on the perceptions of both those whose activities it directly governs and the broader group of actors on whose behalf it purports to govern).

101. For discussions of how internet standards can reflect political considerations, see generally Tarleton Gillespie, *Engineering a Principle: 'End-to-End' in the Design of the Internet*, 36 SOC. STUD. SCI. 427 (2006); LAURA DENARDIS, PROTOCOL POLITICS: THE GLOBALIZATION OF INTERNET GOVERNANCE (2009); Ian Brown, David D. Clark & Dirk Trossen, *Should Specific Values Be Embedded in the Internet Architecture?*, PROC. RE-ARCHITECTING INTERNET WORKSHOP, art. no. 10 (2010). For a discussion specifically focused on the political and normative dimensions of AI standards and their diffusion, see generally Alicia Solow-Niederman, *Can AI Standards Have Politics?*, 71 UCLA L. REV. DISC. 230 (2023).

102. Michael Veale & Frederik Zuiderveen Borgesius, *Demystifying the Draft EU Artificial Intelligence Act*, 4 COMPUT. L. REV. INT'L 97, 105 (2021).

A large firm, for instance, may seek to influence a particular standard's design to raise the costs of smaller rivals and thus further entrench its market position.[103] Intellectual property abuse has also historically been a significant concern in technical standard setting, manifesting through practices such as patent holdup, patent ambush, or royalty stacking, all of which can undermine the accessibility and adoption of standards.[104] Sometimes the problem may be as simple as industry stakeholders steering the design of a standard towards choices that privilege their own interests over public ones, insofar as the two diverge. AI's technical complexity only exacerbates these risks, as less technically sophisticated participants may struggle to detect the ways in which certain design choices could advance narrow industry priorities at the expense of broader societal ones.[105]

Technical knowledge is not the only area of asymmetry that may exist among participants. Disparities in resources for influencing the standards process can be equally problematic. Even when an SDO is nominally "open" and imposes no formal barriers to participation, meaningful engagement requires substantial time and financial investment. For civil society organizations and other non-industry stakeholders, this can serve as a de facto barrier that either severely limits their influence or excludes them from the process altogether.[106] Meanwhile, it is common for larger corporate

103. *See* Olia Kanevskaia, *Governance of ICT Standardization: Due Process in Technocratic Decision-Making*, 45 N.C. J. INT'L L. 549, 551 (2020) ("ICT standards may sometimes result in economic or administrative burdens by pushing up compliance costs for companies."); James J. Anton & Dennis A. Yao, *Standard-Setting Consortia, Antitrust, and High-Technology Industries*, 64 ANTITRUST L.J. 247, 249–50 (1995) (identifying firms' incentives to manipulate standards to raise rivals' costs despite there being "no technical rationale for creating such a disadvantage"); *see also* BREYER, *supra* note 48, at 115 ("While individual standards may not raise barriers significantly, a series of several standards . . . may well raise costs to the point where new firms will find it difficult to assemble sufficient capital to enter.").

104. *See, e.g.*, Mark A. Lemley & Carl Shapiro, *Patent Holdup and Royalty Stacking*, 85 TEX. L. REV. 1991 (2007); Janice M. Mueller, *Patent Misuse Through the Capture of Industry Standards*, 17 BERKELEY TECH. L.J. 623 (2002); Joseph Farrell, John Hayes, Carl Shapiro & Theresa Sullivan, *Standard Setting, Patents, and Hold-Up*, 74 ANTITRUST L.J. 603 (2007).

105. A similar phenomenon has also been observed in the administrative law context. *See* Wendy E. Wagner, *Administrative Law, Filter Failure, and Information Capture*, 59 DUKE L.J. 1321, 1333 (2010) ("[T]he ability to gain control of the rulemaking process through the use of excessive information may even be turned into a strategic advantage. Using technical terms and frames of reference that require a high level of background information and technical expertise, and relying heavily on 'particularized knowledge and specialized conventions,' these fully engaged stakeholders can deliberately hijack the proceedings.").

106. ADA LOVELACE INST., INCLUSIVE AI GOVERNANCE: CIVIL SOCIETY PARTICIPATION IN STANDARDS DEVELOPMENT 26 (Mar. 2023) (identifying the biggest barriers to civil society participation as "restrictive eligibility criteria . . . burdensome time

stakeholders to employ specialists who focus exclusively on standards work, permitting them to monitor and respond to every incoming standards proposal as well as to submit detailed proposals of their own at high volume.[107] This organizational capacity allows them to punch above their weight in the standards development process, which is often well worth the investment due to the financial stakes involved.[108] The upshot of punching above one's weight, however, is that it contributes to the perception that a particular SDO is captured by corporate interests, undermining the legitimacy of both the venue itself and any standards it produces.

That said, because this is a *comparative* second-best analysis, it is important to recognize that many of the challenges described above are not exclusive to standards setting or to private governance more generally. Resource disparities between different interest groups inevitably lead to differences in ability to shape governance outcomes regardless of the institutional arrangement. In the context of more state-centric modalities, these same dynamics express themselves in different forms, whether that be large firms dominating the notice and comment process in informal agency rulemaking or leveraging deep lobbying networks to track and influence legislative developments across all fifty states.[109] Indeed, a vast body of legal literature on regulatory capture documents how well-resourced actors often shape public governance mechanisms in ways that advantage their interests.[110]

To be sure, one could argue that corporate influence may be more pronounced in the context of private standards because commercial actors play a *direct* role in determining outcomes rather than merely influencing the process from the outside. At the same time, there is reason to believe the threat of corporate dominance may be somewhat exaggerated. The commercial actors involved in the standards process are not a monolith with perfectly aligned interests. The AI ecosystem consists of many types of firms operating at different layers of the AI stack (or different stages of the AI supply chain,

commitments, an inability to navigate complicated standardisation processes, industry dominance and a lack of awareness and interest.").

107. *See* BÜTHE & MATTLI, *supra* note 50, at 47.

108. *See id.* at 47–48.

109. *See, e.g.*, Wagner, *supra* note 105, at 1336–37; Elizabeth Warren, *Corporate Capture of the Rulemaking Process*, REGUL. REV. (June 14, 2016), https://www.theregreview.org/2016/06/14/warren-corporate-capture-of-the-rulemaking-process/ (citing an EPA study which found that "industry groups engaged in 170 times more informal communications with EPA than public interest players.").

110. The foundational work that gave rise to the theory of regulatory capture (despite not mentioning it by name) and that influenced decades of scholarship on the relationship between regulation and industry is George J. Stigler, *The Theory of Economic Regulation*, 2 BELL J. ECON. & MGMT. SCI. 3 (1971).

depending on how one conceptualizes it). This includes hardware vendors (e.g., GPU manufacturers), cloud infrastructure providers, model developers, and developers of AI "wrapper" applications—each with distinct and sometimes competing priorities. The heterogeneity of commercial interests can prevent industry stakeholders from speaking with a single voice and serve as a check against any one particular group dominating the entire process.[111] Of course, this alone does not prevent *non-industry* stakeholders from having their voices drowned out, nor does it guarantee that outcomes will align perfectly with public interests. Ensuring balanced representation of interests will also depend heavily on the institutional configuration of the standard-setting venue.

Where SDOs and other technology-focused private governance institutions have historically succeeded in maintaining their legitimacy, it has been due in large part to the openness—both in the participatory and informational sense—of their structures and processes.[112] Within SDOs specifically, openness typically manifests through transparent consensus-based mechanisms that allow for *meaningful* input and engagement from diverse interest groups (i.e., allowing participants to overcome any de facto barriers), thereby preventing any single constituency from dominating the process.[113] This helps ensure that standards are seen as the product of fair and democratic processes that are responsive to the concerns and interests of a broad range of stakeholders.

Thoughtful design of an SDO's structures and processes can go a long way to enhance openness while helping to mitigate concerns about corporate capture. This is likely to include substantive policies and procedural safeguards

111. *Cf.* David D. Clark, John Wroclawski, Karen R. Sollins & Robert Braden, *Tussle in Cyberspace: Defining Tomorrow's Internet*, 13 IEEE/ACM TRANSACTIONS NETWORKING 462 (2005) (arguing that the internet architecture reflects ongoing economic and political "tussles" among stakeholders with competing interests, and that system design should anticipate and balance such tensions).

112. *See* Mulligan & Bamberger, *supra* note 4, at 771 (explaining that past multistakeholder processes "have consistently focused on totems of participation . . . and transparency . . . for procedural legitimacy."); Kanevskaia, *supra* note 103, at 557 (maintaining that the standards produced by industry consortia who limit stakeholder participation and afford limited due process are more likely to lack legitimacy); *see also, e.g.*, A. Michael Froomkin, *Habermas@discourse.net: Toward a Critical Theory of Cyberspace*, 116 HARV. L. REV. 749, 798–805 (2003) (analyzing the IETF's governance model and concluding its open, participatory structure and processes satisfy the procedural requirements for legitimacy under Habermasian discourse theory).

113. *See, e.g.*, AYRES & BRAITHWAITE, *supra* note 79, at 57 (arguing that the concept of "contestability," which can serve as a countervailing force against the threat of regulatory capture, demands transparency and open access to information so that all interested groups can participate meaningfully).

in areas such as standards deliberation, consensus formation, representation in SDO leadership positions, and conflict of interest management (e.g., ex ante disclosure and licensing obligations for IP rights). Designing a system that successfully balances technical expertise, diverse stakeholder interests, practical implementability, and—as the following Section discusses—speed is an undeniably difficult task. The Internet Corporation for Assigned Names and Numbers (ICANN), the multistakeholder body that develops policies governing the technical management and distribution of internet name and number resources, illustrates this point. ICANN has created a complex institutional architecture aimed at ensuring balanced representation of stakeholder interests, including internal policymaking bodies organized and subdivided by stakeholder type,[114] a ruling Board of Directors composed of members drawn from diverse constituencies,[115] and various transparency, accountability, and review mechanisms.[116] This intricate structure notwithstanding, it is not uncommon to hear complaints about a small group of commercial players—particularly domain registries and registrars—wielding disproportionate influence over ICANN's activities.

Nevertheless, ICANN's continued functioning as a global internet governance body, despite its many imperfections, demonstrates that private governance can succeed even in areas where competing commercial, political, and societal interests exist.[117] For AI governance, the key lies in learning from these experiences and adapting institutional designs to the specific challenges of the field. AI standards bodies can indeed create governance frameworks that harness industry expertise while maintaining broader legitimacy. However, this will likely necessitate, among other things, the careful calibration of consensus thresholds (e.g., requiring consensus *within* each stakeholder group rather than across all participants),[118] the creation of targeted support

114. *See* Bylaws for Internet Corporation for Assigned Names and Numbers arts. III-VI (amended Jan. 9, 2025) [hereinafter ICANN Bylaws].

115. *See id.* arts. VII-VIII.

116. *See id.* arts. III-VI.

117. *See generally* Hortense Jongen & Jan Aart Scholte, *Legitimacy in Multistakeholder Global Governance at ICANN*, 27 GLOB. GOVERNANCE 298, 320 (2021) ("ICANN's experience shows what extent of legitimacy a global multistakeholder arrangement can realize in the early twenty-first century, particularly if it undertakes sustained intensive efforts to build support.").

118. It is crucial to understand that in the context of standards development, "consensus" does not equate to "unanimous consent." While there is no universal bright-line rule for determining when consensus exists, it is typically understood as "general agreement, characterized by the absence of sustained opposition to substantial issues by any important part of the concerned interests and by a process that involves seeking to take into account the views of all parties concerned and to reconcile any conflicting arguments." INT'L ORG. FOR STANDARDIZATION & INT'L ELECTROTECHNICAL COMM'N, *ISO/IEC Guide 2:2004, Standardization and Related Activities—General Vocabulary* ¶ 1.7 (2004). To better illustrate the

mechanisms for underrepresented stakeholders, and the establishment of mechanisms for enhanced transparency and managing conflicts of interest.

C. AGILITY

1. Greater Speed and Adaptability

One of the greatest challenges traditional regulation faces in governing emerging technologies is keeping pace with their rapid development.[119] There are at least two reasons for this. First, regulatory processes are often encumbered by procedural frictions. The rulemaking process typically involves multiple mandatory steps—such as notice and comment periods, impact analyses, and interagency reviews—that can take years to complete.[120] Even after rules are enacted, updating them to account for new developments may require going through the same time-consuming procedures. And that is not even to mention how partisan politics can further exacerbate these delays, making regulatory updates even slower and less responsive.

The second reason is that there is an information lag inherent to the regulatory process. Due to the fact that regulators are typically not "on the ground" where the innovation is happening, they often become aware of new advancements well after they have already occurred. Regulators are neither first-hand observers of what goes on in the R&D labs of major technology companies, nor are they privy to the conversations taking place in the hallways of industry conferences and trade shows. It is only once information about these developments eventually trickles down that they recognize the need to act, putting them perpetually behind the curve.[121]

AI bears all of the same regulatory challenges historically faced by other fast-changing emerging technologies and more, as it is not only evolving

---

implications of this definition, consider a scenario where civil society stakeholders constitute 15% of the total participants in an SDO and a large majority of these stakeholders are in objection to a proposed decision. It is still theoretically possible for "consensus" to be reached provided that virtually all of the remaining 85% of participants support the decision and there has been a genuine effort to consider civil society's objections and resolve any disagreements. Such a scenario could be prevented, however, if an SDO had rules defining different classes of stakeholders and requiring consensus to be present in each of them in order for a decision to be approved.

119. This is sometimes referred to as the "pacing problem." *See* Gary E. Marchant, *The Growing Gap Between Emerging Technologies and the Law*, *in* THE GROWING GAP BETWEEN EMERGING TECHNOLOGIES AND LEGAL-ETHICAL OVERSIGHT: THE PACING PROBLEM 19, 22–23 (Gary E. Marchant, Braden Allenby & Joseph Herkert eds., 2011).

120. *See* Lobel, *supra* note 6, at 390.

121. *See* Gervais, *supra* note 47, at 702 ("The speed of technological development also means that once it enters into force, regulation may, in fact, be outdated.").

quickly but doing so in unpredictable directions.[122] Two recent examples perfectly illustrate how the rapid pace of AI innovation is already straining traditional regulatory approaches.

The first can be seen in the initial draft of the EU AI Act. European regulators spent over a year devising a comprehensive framework built around the assumption that AI systems would be developed to serve individual use cases.[123] This assumption was turned on its head after the public release of ChatGPT, which became one of the most rapidly adopted technologies in history, surpassing one hundred million users within just two months of its introduction.[124] Prior to this point, large language models (LLMs) and other general-purpose AI did not appear to have been on the radar of European regulators despite OpenAI having already released multiple iterations of its GPT model in the preceding years (albeit to much less fanfare).[125] The sudden prominence of ChatGPT forced lawmakers to overhaul their draft legislation to account for general-purpose AI, delaying the regulatory process.[126]

A second example can be found both in the EU AI Act as well as an unsuccessful piece of AI safety legislation at the state level, California's proposed SB 1047. Both measures define "compute thresholds"—specific levels of computational power used during training that trigger regulatory obligations—to determine the responsibilities of AI model developers.[127] In the EU AI Act, models trained using at least 10^{25} floating point operations (FLOPS) of computation are presumed to be classified as "general purpose AI with systemic risk," a category that carries heightened regulatory obligations.[128] Similarly, SB 1047, a since-vetoed bill which aimed to mitigate the most serious

122. *See* Marchant & Gutierrez, *supra* note 90, at 384; *see also* Gervais, *supra* note 47, at 687, 701–02 (highlighting the dangers of regulatory interventions directed at "inchoate technologies" that evolve quickly along unpredictable trajectories).

123. *See* Gian Volpicelli, *ChatGPT Broke the EU Plan to Regulate AI*, POLITICO EU (Mar. 3, 2023), https://www.politico.eu/article/eu-plan-regulate-chatgpt-openai-artificial-intelligence-act/; Natali Helberger & Nicholas Diakopoulos, *ChatGPT and the AI Act*, 12 INTERNET POL'Y REV. 1 (2023), https://doi.org/10.14763/2023.1.1682.

124. Dan Milmo, *ChatGPT Reaches 100 Million Users Two Months After Launch*, GUARDIAN (Feb. 2, 2023), https://www.theguardian.com/technology/2023/feb/02/chatgpt-100-million-users-open-ai-fastest-growing-app.

125. *See* Volpicelli, *supra* note 123.

126. *See id.*

127. *See* Artificial Intelligence Act, *supra* note 13, art. 51(2); S.B. 1047, § 3, 2023–2024 Leg., Reg. Sess. (Cal. 2024) (as enrolled Sep. 3, 2024), https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240SB1047 [hereinafter S.B. 1047].

128. *See* Artificial Intelligence Act, *supra* note 13, arts. 51, 55.

risks posed by the largest and most advanced AI models, would have imposed obligations on developers of models trained using more than 10^{26} FLOPS.[129]

Although both the EU AI Act and SB 1047 provided for mechanisms through which these thresholds could be amended over time, such updates are still subject to formal processes that inevitably introduce delay.[130] Indeed, in just the few months since California Governor Gavin Newsom vetoed SB 1047, major advancements in training techniques have already begun to challenge the assumption that achieving frontier-level capabilities requires such massive computational resources. Most notably, a Chinese hedge fund released an open-source model called DeepSeek R1 that reportedly matched the performance of advanced Western models despite using considerably less training compute.[131] This development led multiple commentators in the AI policy community to point out that, had SB 1047 been signed into law, its compute thresholds would have already been rendered obsolete before they even had a chance to take effect.[132] Beyond DeepSeek, similar risks of obsolescence loom with the rise of techniques like chain-of-thought reasoning and retrieval-augmented generation, which shift much of the computational burden to "inference time" and can produce sharp post-training jumps in model capability without increasing training compute.[133] In short, unforeseen technical developments—far from uncommon in an area as fluid as AI—can quickly frustrate even the most forward-thinking hard law schemes.

By comparison, private standards-based governance offers a nimbler alternative that can better adapt to external changes in the technological landscape. Since it is not *formally* burdened by the same constraints traditional regulation faces, standards bodies can potentially respond more promptly to

129. *See* S.B. 1047, *supra* note 127, § 3; *see also* Letter from Gavin Newsom, Governor of Cal., to Members of the Cal. State Sen. (Sep. 29, 2024), https://www.gov.ca.gov/wp-content/uploads/2024/09/SB-1047-Veto-Message.pdf.

130. *See* Artificial Intelligence Act, *supra* note 13, art. 51(3); S.B. 1047, *supra* note 127, § 3.

131. *See* Caiwei Chen, *How a Top Chinese AI Model Overcame U.S. Sanctions*, MIT TECH. REV. (Jan. 24, 2025), https://www.technologyreview.com/2025/01/24/1110526/china-deepseek-top-ai-despite-sanctions/; Sarosh Nagar & David Eaves, *AI's Efficiency Wars Have Begun: The DeepSeek Shock May Reshape a Global Race*, FOREIGN POL'Y (Feb. 5, 2025), https://foreignpolicy.com/2025/02/05/deep-seek-china-us-artificial-intelligence-ai-arms-race/. *But see* Dylan Patel, AJ Kourabi, Doug O'Laughlin & Reyk Knuhtsen, *DeepSeek Debates: Chinese Leadership on Cost, True Training Cost, Closed Model Margin Impacts*, SEMIANALYSIS (Jan. 31, 2025), https://semianalysis.com/2025/01/31/deepseek-debates/ (contesting widely reported estimates of DeepSeek's training costs).

132. *See, e.g.*, Timothy B. Lee (@binarybits), X, *Imagine how dumb the California legislature would look today if Newsom had signed SB 1047 last fall instead of vetoing it.* (Jan. 27, 2025, at 03:15 PM), https://x.com/binarybits/status/1751710571690037674.

133. *See* Sara Hooker, *On the Limitations of Compute Thresholds as a Governance Strategy*, ARXIV:2407.05694, at 5, 13–14 (July 8, 2024), https://arxiv.org/abs/2407.05694.

evolving technologies and emerging risks. It is worth noting that SDOs do end up recreating some of these constraints anyway, as many of the aforementioned "frictions" are in fact the product of procedural devices that help preserve legitimacy and ensure alignment with certain democratic ideals.[134] Still, private standards bodies have more room for institutional and process innovations, leaving them better equipped to strike an optimal balance between speed and legitimacy.[135] New initiatives and groups can also quickly convene without needing to worry about underlying legal authorities or possible judicial challenges like an administrative agency would.

As for the information lag, many of the private actors who participate in standards development are far less susceptible to it. Given their direct involvement in the ecosystem, these private actors are better positioned to see what is happening on the ground and gain a better sense of the technology's trajectory as it continues to develop. It is difficult to imagine these actors being caught off guard by major technological advancements that they themselves are responsible for. This closer proximity to technological development allows private standards bodies to more quickly identify when existing standards need revision or when entirely new approaches are warranted.

2. Second-Best Considerations: Prematurity and Tension with Legitimacy

Despite the potential for private standards processes to move faster than traditional regulation, there are pitfalls associated with moving *too* quickly and standardizing prematurely. Taking an anticipatory approach to standardization—attempting to get ahead of technological developments by establishing standards before a technology can be fully built and tested—requires that developers design standards without seeing what actually works in practice, forcing them to make assumptions about future needs and technological conditions.[136] Admittedly, this anticipatory approach may be seen as attractive for developing AI safety standards because it would appear better positioned to proactively address risks of harm before they materialize. According to this line of reasoning, permitting experimentation during the early stages and waiting for measures and techniques with a proven track record to emerge may needlessly subject the public to dangers that could

134. *See* Kanevskaia, *supra* note 103, at 554–55 (explaining that transnational private regulators, a category that can include technical SDOs, will often address legitimacy concerns through procedural frameworks rooted in administrative law principles such as due process, transparency, participation, and review).

135. *See* Lobel, *supra* note 6, at 390–91.

136. This challenge is exacerbated when the object being standardized is a component of a larger, complex system and thus has interdependencies (even if minimal) with other parts. *See* Christopher S. Yoo, *Modularity Theory and Internet Regulation*, 2016 U. ILL. L. REV. 1, 8–9.

otherwise be avoided by standardizing up front. However, standardizing prematurely can result in unproven standards that either fail to gain adoption, or even if they do, prove inefficient and ineffective—giving only the illusion of safety or security.[137]

The difficulty of developing effective standards through an anticipatory approach was one of the main lessons of the OSI-IETF internet standards war of the late 1980s and early 1990s. ISO's Open Systems Interconnect (OSI) architecture, which had the backing of many governments around the world, was carefully planned out in advance through a forward-thinking, bureaucratic process.[138] However, ISO's proactive standardization resulted in an over-engineered architecture that proved far more complex than was practical; the designers of OSI failed to account for the type of network capabilities for which there was legitimate demand and that had been proven to work in practice.[139] Meanwhile, the IETF's "rough consensus and running code" approach, which prioritized judging functional, real-world implementations of a prospective standard on its technical merits before approving it, produced the suite of protocols that ultimately prevailed.[140] Hence, even though SDOs may feel pressure to produce standards quickly, they must carefully consider the tradeoffs between speed and effectiveness.[141]

A closely related challenge and one that this Article briefly alluded to earlier is that moving fast often comes at the expense of legitimacy. Designing the structure and rules of an SDO to prioritize speed can compromise the perceived inclusiveness and impartiality of the standards process. For instance,

137. *See* NIST AI Standards Plan, *supra* note 29, at 5 ("Conversely, a standard that would attempt to get ahead of the underpinning science and engineering may be built on less rigorous technical foundations; it may prove unhelpful, counterproductive, or even technically incoherent.").

138. *See* JANET ABBATE, INVENTING THE INTERNET 168 (1999) ("The network standards effort was a departure from ISO's usual practice, in that it represented an attempt to standardize a technology that was still new and had not had a chance to stabilize . . . . But in the case of networks, some ISO members felt that formal standards should be outlined proactively.").

139. *See id.* at 176 (noting that many IT professionals saw OSI as unnecessarily complex and inefficient); *see also* Andrew L. Russell, *'Rough Consensus and Running Code' and the Internet-OSI Standards War*, 28 IEEE ANNALS HIST. COMPUTING 48, 53–54 (2006) (noting the internet engineers tended to view the OSI's design approach as "out of touch with existing networks and computers.").

140. *See* Russell, *supra* note 139, at 55.

141. A recent report on AI security standards published by the Alan Turing Institute's Centre for Emerging Technology and Security found that many SDOs who typically err on the side of technical maturity have reportedly been "facing pressure to standardize now." *See* Rosamund Powell, Sam Stockwell, Nalanda Sharadjaya & Hugh Boyes, *Towards Secure AI: How Far Can International Standards Take Us?*, CTR. FOR EMERGING TECH. & SEC. 31, 45 (Mar. 2024), https://cetas.turing.ac.uk/publications/towards-secure-ai.

providing for an override mechanism that allows quick resolution when disagreements arise—such as resorting to majority voting instead of continuing attempts at consensus-building—may expedite the process but leave stakeholders in the minority feeling marginalized.[142] Similarly, imposing narrow time windows where participants can provide input or voice objections to standards proposals may streamline standardization but at the cost of shutting out certain stakeholders, particularly those who lack the resources to respond promptly.[143] These process design choices, while facilitating agility, can thus undermine the very legitimacy that private standards-based governance requires to remain viable.

Conversely, the same participatory openness that lends legitimacy to SDOs can also hinder their progress by dramatically prolonging the process. As the number of participants grows, so too does the complexity of coordinating their input and resolving their often-conflicting interests. Discussions become unwieldy, decision-making more cumbersome, and consensus increasingly difficult to achieve. Notably, one of the core assumptions underlying Tiebout's foot-voting model is that communities have an optimal size; when this optimum is exceeded, a community can no longer provide local public goods at the same level and cost-efficiency that attracted residents in the first place.[144] Similarly, there comes a threshold beyond which adding more participants to an SDO does little to add to legitimacy and only contributes to crowdedness and gridlock, potentially negating one of the key advantages of standards-based governance. This tension adds an additional layer of complexity to the institutional and process design puzzle: SDOs must not only consider how to balance competing interests but also how to weigh speed and legitimacy when determining who gets a seat at the table and how decisions in the standardization process are made. This itself does not render private standards-based governance unworkable or inferior to alternatives—only demanding of careful institutional design.

D. SCALE

1. Better Positioning to Scale Across Borders

A final key advantage that private standards-based governance maintains over traditional regulation is its ability to facilitate coordination on a global scale. Whereas traditional regulation is territorially delimited—legal commands

142. *See id.* at 32 (reporting that interviewed SDO participants suggested that agility could be increased by "moving away from consensus-based agreements and towards majority-based voting and introducing streamlined approaches to submitting comments on draft standards.").

143. *Id.*

144. *See* Tiebout, *supra* note 54, at 419.

issued by sovereigns typically have no effect outside their jurisdiction—private standards more easily transcend borders. When standards are developed through open, multistakeholder arrangements and the final publications are made widely and freely accessible, they allow for broad participation and adoption from stakeholders regardless of their geographic location. In turn, this can help establish common or interoperable governance frameworks that extend across heterogenous legal systems and market contexts.

Global coordination may be less imperative for AI than for networked technologies and industries whose value depends heavily on widespread compatibility and interconnection. Take the internet, for instance: in order to function as a universal communication infrastructure—a single network that connects thousands of smaller networks from around the world—there must be some degree of global coordination around its core technical protocols as well as its naming and numbering systems.[145] The basic functionality of AI systems, by contrast, is not nearly as dependent on global uniformity: discrete AI systems can be built to conform to certain requirements in one jurisdiction without necessarily affecting those in other jurisdictions. That said, transnational cooperation around AI standards still offers significant advantages. Globally coordinated standards allow firms to develop a single AI system aligned with a widely accepted framework rather than incur the expensive burden of building multiple systems tailored to different local AI ecosystems. These standards can thus facilitate cross-border commerce and improve the ease of doing business for multinational firms that might otherwise face a complex patchwork of conflicting national requirements. They also support knowledge transfer between regions, defining globally recognized best practices that can help developing countries expand their domestic AI capacity and participate more meaningfully in the broader AI economy.

Governments, of course, have their own means of achieving cross-border regulatory harmonization, primarily through supranational organizations and agreements whereby member states develop and then independently enact uniform regulations. In fact, this approach would arguably be more effective at promoting global harmonization than private standards. Whereas the success of the latter is contingent on the absence of conflicting national regulations that would effectively preempt private standards, a multilateral regime aimed at regulatory harmonization would provide a clear roadmap for eliminating conflicting national regulations. However, this approach also has several drawbacks.

145. *See* LAURA DENARDIS, THE GLOBAL WAR FOR INTERNET GOVERNANCE 16–18 (2014).

First, it inherits many of the same weaknesses of national regulation—it is slow, has a top-down orientation, and entrusts policymaking to government representatives who are less likely to possess sufficient technical expertise. Second, it introduces several new complexities, such as conflicts between national values, divergent risk tolerances, and geopolitical rivalries. Multilateral talks around digital public policy issues such as privacy and international data flows have persistently stalled due to fundamental disagreements between major powers, giving little reason to expect that AI regulation would fare better.[146] Ultimately, if meaningful global coordination on AI governance is to be achieved, private standards-based governance represents a much more viable path forward.

2. Second-Best Considerations: Geopolitical Competition

Insofar as cooperation around AI standards does take on a transnational character, it tends to invite state involvement even when standards venues are nominally private. Though governments may not directly control the process, the potential political, economic, and social stakes give them strong incentives to try to influence their development.[147] Countries further recognize that leadership in setting AI standards can yield strategic advantages, conferring soft power, national prestige, and other economic advantages for domestic firms. Larger world powers may be tempted to influence private standards development if for no other reason than to prevent their adversaries from doing the same. If AI is indeed the transformative technology that many believe it to be, concerns will inevitably arise about rival states writing the rules

146. Because cross-border data flows have become an integral part of modern international commerce, attempts at multilateral regulatory cooperation have largely taken place within the context of negotiations over binding trade rules. *See, e.g.*, Anupam Chander & Paul M. Schwartz, *Privacy and/or Trade*, 90 U. CHI. L. REV. 49, 56–60, 65–69 (2023) (recalling the efforts to address transnational privacy issues during the Uruguay Round of negotiations that led to the creation of the WTO and how, rather than find a permanent solution that resolved underlying disagreements, these issues were ultimately "bracketed" through the inclusion of an open-ended GATS exception under which Members could justify trade-restrictive domestic privacy regulations). The most recent opportunity to make meaningful progress—plurilateral negotiations at the World Trade Organization under the Joint Statement Initiative on e-commerce—collapsed in 2023 following persistent disagreements among major economies and the abrupt withdrawal of the United States. *See* Alex Mueller, *One Step Forward, Two Steps Back: The United States' New Direction on Digital Trade*, 26 MINN. J.L. SCI. & TECH. 116 (2025).

147. *See* Hadrien Pouget, *What Will the Role of Standards Be in AI Governance?*, ADA LOVELACE INST. (Apr. 5, 2023), https://www.adalovelaceinstitute.org/blog/role-of-standards-in-ai-governance/ ("Since standards can carry some regulatory influence and relate to issues that matter to governments, governments are invested in their content and disagreements can arise.").

for AI to align with their preferences.[148] Traces of this can already be see in the United States, where AI development is now commonly framed as a race against China to determine whose vision for the technology prevails.[149]

The problem here is that the indirect involvement of states and infiltration of geopolitical rivalry into the standards arena threatens the integrity and effectiveness of the standards development process. It provides a new source of conflict and disagreement that can impede progress and significantly prolong standardization timelines. Similarly, distortions in the standards process arise when a country attempts to influence standards by placing its finger on the scale, potentially leading to suboptimal outcomes such as the adoption of technically inferior standards. There is already limited evidence of such scale tipping activity taking place at private SDOs. For example, participants at various technical SDOs have reported observing highly coordinated behavior by representatives of Chinese firms.[150] The *Wall Street Journal* even reported that, during a leadership election at the 3rd Generation Partnership Project (3GPP) that involved a candidate from Huawei, representatives from other Chinese companies were expected to capture proof that they cast their ballots for the preferred candidate.[151]

148. *See, e.g.*, Katherine Golden, *If the US and EU Don't Set AI Standards, China Will First, Say Gina Raimondo and Margrethe Vestager*, ATL. COUNCIL (Jan. 31, 2024), https://www.atlanticcouncil.org/blogs/new-atlanticist/if-the-us-and-eu-dont-set-ai-standards-china-will-first-say-gina-raimondo-and-margrethe-vestager/ (quoting former U.S. Commerce Secretary Gina Raimondo as warning that "if the US and EU don't show up [to set AI standards], China will, [and] autocracies will.").

149. *See* R. David Edelman, Diana Fu, Ryan Hass, Patricia M. Kim, Ying Lin, Yingyi Ma, Michael E. O'Hanlon, Melanie W. Sisson, Elham Tabassi & Nicol Turner Lee, *How Will AI Influence US-China Relations in the Next 5 Years?*, BROOKINGS INST. (June 18, 2025), https://www.brookings.edu/articles/how-will-ai-influence-us-china-relations-in-the-next-5-years/ (explaining how voices ranging from OpenAI CEO Sam Altman to former National Security Advisor Jake Sullivan have adopted this arms-race-with-China framing); *see also* Alexandra Alper & Jody Godoy, *AI Execs Say U.S. Must Increase Exports, Improve Infrastructure to Beat China*, REUTERS (May 8, 2025), https://www.reuters.com/world/us/us-ai-execs-give-congress-policy-wishlist-beating-china-2025-05-08/ (quoting Microsoft President Brad Smith as saying "the number-one factor that will define whether the U.S. or China wins this race is whose technology is most broadly adopted in the rest of the world.").

150. Emily de la Bruyère, *Setting the Standards: Locking in China's Technological Influence*, *in* CHINA'S DIGITAL AMBITIONS: A GLOBAL STRATEGY TO SUPPLANT THE LIBERAL ORDER 49, 57 (Nat'l Bureau Asian Rsch., NBR Special Rep. No. 97, Emily de la Bruyère et al. eds., 2022); Daniel R. Russel & Blake H. Berger, *Stacking the Deck: China's Influence in International Technology Standards Setting*, ASIA SOC'Y POL'Y INST. 12 (2021), https://asiasociety.org/sites/default/files/2021-11/ASPI_StacktheDeckreport_final.pdf.

151. Valentina Pop, Sha Hua & Daniel Michaels, *From Lightbulbs to 5G, China Battles West for Control of Vital Technology Standards*, WALL ST. J. (Feb. 8, 2021), https://www.wsj.com/articles/from-lightbulbs-to-5g-china-battles-west-for-control-of-vital-technology-standards-11612722698.

It is important to keep in mind, however, that these potential challenges around indirect government influence are far less severe than those that would arise in a more state-centric alternative where governments formally play a controlling role in setting transnational standards. Furthermore, just like many of the other challenges discussed throughout the Article, institutional design has an important role to play. Clear rules for achieving consensus, transparency and oversight mechanisms, and other procedural safeguards can help mitigate the risks of state interference. One specific option that SDOs may want to consider is permitting governments to participate in a limited advisory role—similar to ICANN's Government Advisory Committee (GAC)—where government representatives would have an opportunity to voice concerns over proposed standards without directly influencing consensus-based decision-making.[152] By providing a structured and transparent channel for governments to engage with SDOs, it could reduce the likelihood that they attempt to exert influence through indirect means. Of course, this is just an example, but it illustrates the type of design choices available to SDOs as they work to preserve the advantages of private standards-based governance while minimizing its vulnerabilities.

IV. CONCLUSION: THE PATH FORWARD

While no model for governing AI is perfect, private standards-based governance represents the most viable and compelling path forward. When implemented effectively, standards offer a powerful means of governing AI that can embed measurable expectations and constraints directly into the design, deployment, and oversight of AI systems. This approach offers significant advantages over traditional regulation, which often struggles under the weight of centralized authority, limited technical capacity, and procedural rigidity. By contrast, private standards-based governance stands to benefit from a bottom-up, multistakeholder architecture that enables experimentation and faster iteration, draws on deep field-level expertise and experience, and scales more readily across national borders. Its advantages lie not in its flawlessness but rather in its ability to navigate the realities of governing a complex, fast-moving, emerging technology. That said, this conclusion should

152. *See* ICANN Bylaws, *supra* note 114, art. XII, § 2(a). However, it should also be noted that the GAC has earned a fair share of criticism for the way it pulls an extraordinary group of actors (national governments) into the institutional fold as if it were any other stakeholder group, muddying the waters of power and authority over ICANN's activities. *See, e.g.*, Jonathan Weinberg, *Governments, Privatization, and "Privatization": ICANN and the GAC*, 18 MICH. TELECOMM. & TECH. L. REV. 189 (2011); MILTON MUELLER, NETWORKS AND STATES: THE GLOBAL POLITICS OF INTERNET GOVERNANCE 242–44 (William J. Drake & Ernest J. Wilson III eds., 2010).

not be mistaken for complacency, as there is still substantial work that remains if these advantages are to be realized.

Although many AI standardization initiatives are already underway, the task of building a robust standards-based regime is nowhere near complete. This reality is perhaps lost on some. In the United States, for example, most of the conversation around AI standards revolves heavily around NIST's AI RMF, which has led to the tendency to treat this framework as if it were intended as an exhaustive, turnkey solution to all of the AI governance challenges we currently face. To be sure, the AI RMF plays a valuable role in helping establish a common language for thinking about the organizational dimensions of AI governance. And NIST's supporting materials—such as its context-specific "profiles" and implementation "playbooks"—have made the framework more practical and usable for a variety of stakeholders.[153] But this body of work is, at most, only a first step. It does not (and was never meant to) address many of the technical dimensions of AI governance and cannot substitute for the full suite of standards needed to manage AI risk at scale across sectors and system types.

As AI standards work continues to unfold, several questions related to both the substantive and procedural aspects of AI standardization remain unanswered and will need to be confronted if this work is to fulfill the governance ambitions outlined throughout the Article. This includes the many questions related to SDO structure and process design that have been posed throughout this Article, such as how to mitigate the risks of industry capture or how to manage the tension between speed and legitimacy. There is also the "timing" question discussed in the previous Part: how to balance the mounting pressure to develop AI standards now with the practical reality that higher-quality standards often require greater technical maturity and iterative learning. Beyond these challenges, there are several more intricate questions about the appropriate scope and structure of standards themselves. A well-developed understanding of when and how standards should intervene across different stages of the AI system lifecycle remains lacking. For example, does pre-deployment testing and validation require different benchmarks and procedures than post-deployment? How many different modes of testing or validation are needed within a given stage, and do they require discrete standards?

153. *See, e.g.*, NAT'L INST. OF STANDARDS & TECH., AI RISK MANAGEMENT FRAMEWORK PLAYBOOK, *supra* note 38; NAT'L INST. OF STANDARDS & TECH., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK: GENERATIVE ARTIFICIAL INTELLIGENCE PROFILE (NIST AI 600-1, 2024), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf.

It would be a mistake to assume that all of these challenges and questions will inevitably resolve themselves with time. This Article does *not* argue that private standards-based governance is inherently self-correcting and that no further actions are necessary. Instead, as repeatedly underscored, the success of private standards-based governance will be heavily contingent on institutional design, or more specifically, configuring SDOs and their processes to make them more effective, inclusive, and legitimate. Sadly, several of the existing SDOs heavily involved in ongoing AI standards development leave much to be desired in this regard. Take ISO/IEC, for instance, which represents the earliest and perhaps most prominent AI standardization initiative to date. Its participation model requires stakeholders to engage indirectly through nationally designated standards bodies—entities that set their own rules for membership, engagement, and decision-making.[154] This structure leaves no formal avenues for direct participation by independent experts or civil society organizations (especially those who are not members of their respective national bodies), often limiting meaningful access to the standard-setting process.[155] Compounding these challenges, the resulting standards are typically locked behind paywalls and cost several hundred dollars to purchase, making it difficult for many stakeholders to access the very specifications intended to guide responsible AI development.[156] These major shortcomings in accessibility and transparency should give all stakeholders pause about whether current standardization efforts are on the right trajectory, and just as importantly, prompt renewed attention to their role in shaping these processes from the inside.

At the same time, responsibility for the success of private standards-based governance cannot fall solely on private actors. Traditional policymakers also have a role to play—not as top-down regulators, but as facilitators of effective and inclusive governance. In addition to protecting the integrity of the standards process by enforcing commitments and policing unfair or anticompetitive conduct within SDOs, they can serve as benign but ever-present backstops, strategically wielding the threat of formal regulation to steer private actors toward greater cooperation and a willingness to adopt more

154. *See* Kanevskaia, *supra* note 103, at 613.

155. *See id.* (noting that consumer and civil society groups can apply for "liaison" status in ISO/IEC technical committees, but the limited voting rights means they must resort to lobbying if they want to influence outcomes).

156. For example, ISO/IEC 42001, the foundational AI management standard, is priced at 199 Swiss francs (approximately 250 USD as of July 2025). *See* ISO, *ISO/IEC 42001:2023 Information Technology — Artificial Intelligence — Management System*, https://www.iso.org/standard/42001.

rigorous standards.[157] Policymakers should also consider the recommendations of those who have advocated for correcting imbalances by investing in the capacities of marginalized stakeholders, such as through stipends, technical assistance, or other focused interventions that reduce resource and knowledge gaps that influence standards process outcomes.[158]

In the end, the question is not whether standards will govern AI; indeed, they are already starting to do so.[159] The real question is whether policymakers and stakeholders will recognize their emergence for what it is—not just an inevitability but an opportunity—and respond with the seriousness it demands. For standards to serve as a legitimate and sustainable form of governance, there must be active investment in their design and the building of institutional foundations necessary for accountability, inclusivity, and public trust. Private standards-based governance offers great promise: it is agile, expert-driven, and globally scalable in ways that traditional regulation is usually not. But there is nothing inevitable about realizing this promise. Whether it becomes a tool for public good or merely the manifestation of particular interests will depend entirely on the actions taken going forward.

157. *Cf.* AYRES & BRAITHWAITE, *supra* note 79, at 38–41 (discussing the idea of the "benign big gun," wherein regulators can "speak softly" while maintaining a credible willingness to escalate up a regulatory pyramid toward more coercive strategies if voluntary cooperation fails).

158. *See* Margot E. Kaminski, *Voices In, Voices Out: Impacted Stakeholders and the Governance of AI*, 71 UCLA L. REV. DISC. 176, 194–95 (2024) (arguing that support for capacity-building efforts is necessary if stakeholders are going to meaningfully participate in AI governance); Mulligan & Bamberger, *supra* note 4, at 775 ("Meaningful participation in design debates further requires resources and strategies to bolster the uneven technological expertise among stakeholders.").

159. *See supra* text accompanying notes 33–39.

STUDENTS, POWER, AND TECHNOLOGY

Mona Sloane[†]*, Ella Duus*[††]*, and Bertrall Ross*[†††]

ABSTRACT

Digital technology and artificial intelligence (AI) systems have become deeply integrated into the operations and pedagogies of universities. Despite being pervasive and creating digital footprints for students that may last a lifetime, these technologies remain ungoverned by its key stakeholder group: students. This development coincides with a steady decline in student power in university administration as evidenced by recently changed state laws, structurally hindering student participation in technology governance.

TABLE OF CONTENTS

I. INTRODUCTION

The 2024 arrival of generative artificial intelligence (AI) systems, such as OpenAI's ChatGPT, Google's Gemini, or Anthropic's Claude, on U.S. campuses threw administrators and faculty into crisis: suddenly, students could use this technology to complete almost all course assignments, undermining pedagogical approaches and assessment procedures that had been built over decades. In response, most universities hastily installed working groups or

DOI: https://doi.org/10.15779/Z38JS9HB12

† Assistant Professor of Data Science and Media Studies, University of Virginia; Faculty Co-Lead Digital Technology for Democracy Lab, University of Virginia; Director, Sloane Lab.

†† Graduate Student, Frank Batten School of Leadership and Public Policy, University of Virginia; Research Assistant, Sloane Lab.

††† Professor of Law, University of California, Berkeley, School of Law; Director, Edley Center on Law & Democracy.

committees tasked with creating usage and policy recommendations, typically focused on ensuring plagiarism could be curbed, following paternalistic and punitive approaches to this technology. Additionally, most of these committees and working groups were composed of faculty and administrators, centering their concerns, rather than taking student perspectives into account.

While generative AI is often seen as the cause of fundamental changes in higher education learning, it brings into sharp relief a more important and structural issue that has been brewing since the 1960s: the systematic decline of student power on public university boards and participation in university governance, paired with the rapid growth of digital technology in student life and learning.

Although students constitute a key stakeholder group on campuses, they are increasingly excluded from exercising agency in the university context, especially when it comes to the digital technologies they are forced to use to participate in campus life, access essential services, and engage meaningfully in the classroom. Increasingly, these technologies involve privacy-invasive data collection and harbor the risk of creating student data footprints that are consumed by AI systems in a competitive AI training data market in which data is an increasingly valuable commodity. An absence of student governance of digital technologies poses the risk of coercing students into being profiled, potentially affecting how future AI systems assess them in high-stakes situations, such as pursuing a job, loan, work visa, or insurance.

In this Article, we describe and analyze the interlocking trends of growing pervasive digital technology on campuses and rapidly declining student participation in university governance, including technology governance. We contextualize these entangled developments with the history of student participation in university administration and shared governance. We argue that justifications for student exclusion or marginalization from university governance are superficial and cause universities to miss out on leveraging students' comparative advantage as early technology adopters and active users while violating students' agency and their distinct need for intellectual privacy. We build on decision-making theory to suggest that excluding students from governance will lead to worse technology procurement, innovation, and regulatory outcomes and propose the Student Technology Council (STC) as a new model for addressing this risk.

In Part II, we describe how data-harvesting technologies have become integral to student life inside and outside of the classroom. We also show how universities have ceded power over technology governance to third-party vendors, changing how students interact with the institution and often violating their privacy expectations without providing proof of the added benefit of the newly procured digital tools.

In Part III, we contrast these findings with an analysis of current technology governance approaches at universities, most of which grew out of corporate information technology (IT) governance approaches that are decentralized today. These IT governance approaches, which are administrator- and faculty-led, rely on incorrect assumptions and incomplete assessments about student technology use and privacy expectations.

In Part IV, we converge on the idea of the marginalized student in university governance, describing how the current state of student exclusion from core governance work and technology decisions is at odds with the historically powerful role students held at universities from the beginning. By surveying the history of student power, we find that students have historically gained structural inclusion in governance after mobilizing as a distinct political group, such as in protests. In the United States, student power and governance participation increased after protests in the 1960s. However, that power has since been rolled back, particularly through state laws governing the composition of public university boards and changes to the student-university relationship, all toward a model of contracting parties whereby students are positioned as buyers of an educational product.

In Part V, we argue that there is no legal basis for excluding students from technology governance and that decision-making theory holds that continuing on this trajectory will cause worse technology and governance outcomes for universities writ large. We conclude the paper by suggesting the installment of STCs, elected bodies of students that advise administrators, faculty, and departmental and cross-university committees on issues and questions pertaining to technology procurement, innovation, and governance by issuing recommendations and opinions and by engaging the student body directly in technology conversations. The STC would give students a formal voice in shaping university technology decisions and ensure their expertise and well-being are prioritized. Ultimately, such a participatory form of governance can leverage students' backgrounds, lived technology experiences, and specific privacy expectations to enable better-informed and legitimate technology decisions on technology procurement, innovation, and governance, positioning students as co-governors rather than marginalizing their expertise and agency.

II. THE ROLE OF DIGITAL TECHNOLOGY IN STUDENT LIFE

Across American higher education institutions, data-intensive technologies are embedded into every aspect of campus life and learning, affording third-party technology vendors outsized power and control over the intimate lives of students, fundamentally changing how they interact with the

institution, and posing potential privacy risks with oftentimes nebulous benefits. University Wi-Fi networks, internet-connected devices, and sensors feed constant streams of data into predictive analytics platforms, AI-powered software, and learning management systems (LMSs). Administrators and product vendors promote these technologies as mechanisms to optimize resource use, personalize instruction, and improve educational quality. Universities accordingly collect increasingly granular student information and procure systems to make decisions and interventions based on that data. For example, some schools have gone so far as pinging students' phone locations to track class attendance[1] or requiring students to wear Fitbits for a physical education course.[2] Indeed, universities utilize technology both outside and inside the classroom.

Even before students set foot on campus, higher education institutions increasingly leverage technology to shape the student experience. For example, universities create profiles on prospective students based on their demographics, financial background, application, and interest forms from standardized tests.[3] They then use predictive analytics to calculate a student's "demonstrated interest" and the amount of financial aid needed to attract a student.[4] Once on campus, algorithms help curate the emails students receive, the courses recommended to them, the real-time alerts sent to advisors monitoring their academic performance and well-being, and the extracurriculars, scholarships, and opportunities that are shared with them.[5] Even when students leave the university, their data is still used. Technologies like Degree Compass at Austin Peay State University permanently retain historical grade data to compare with current student transcripts and recommend classes.[6] These systems optimize for metrics that improve

1. Machaella Reisman, *University Use of Big Data Surveillance and Student Privacy*, 48 FLA. ST. U. L. REV. 559, 564–65 (2021).

2. Jonathan Root, *How Fitbit Helps a Conservative Evangelical College Monitor Students' Bodies for Christ*, RELIGION DISPATCHES (Mar. 10, 2016), https://religiondispatches.org/2016/03/10/how-fitbit-helps-conservative-evangelical-college-monitor-students-bodies-christ.

3. *See* Victor M.H. Borden & Hamish Coates, *Learning Analytics as a Counterpart to Surveys of Student Experiences*, 2017 NEW DIRECTIONS HIGHER EDUC. 89, 90.

4. Tim Lloyd, *How College Applications Change in the Era of Big Data*, MARKETPLACE (Jan. 14, 2014), https://www.marketplace.org/story/2014/01/14/how-college-applications-change-era-big-data; Maggie McGrath & Matt Schifrin, *The Invisible Force Behind College Admissions*, FORBES (July 30, 2014), https://www.forbes.com/sites/maggiemcgrath/2014/07/30/the-invisible-force-behind-college-admissions/; LINDSAY WEINBERG, SMART UNIVERSITY: STUDENT SURVEILLANCE IN THE DIGITAL AGE 57–59 (2024).

5. WEINBERG, *supra* note 4, at 1, 3, 17–18, 57–59.

6. MANUELA EKOWO & IRIS PALMER, THE PROMISE AND PERIL OF PREDICTIVE ANALYTICS IN HIGHER EDUCATION: A LANDSCAPE ANALYSIS 9 (2016), https://files.eric.ed.gov/fulltext/ED570869.pdf.

colleges' rankings and state performance-based funding, such as students' duration of study and retention rate.[7] They allow universities to predict and manage a more expected future and stable student population, while improving their rankings and maintaining funding.[8]

Within the classroom, students' learning is mediated by technology, too. This includes LMSs, digital dashboards that track student performance, adaptive learning tools, automated feedback systems, and online proctoring services such as Respondus[9] and ProctorU.[10] Faculty are increasingly asked to use LMSs to integrate digital learning materials, assignments, grades, discussion boards, and video and audio recordings. For example, faculty at the City University of New York (CUNY) now require all synchronous and asynchronous online classes to be delivered through the LMSs Brightspace or Blackboard.[11] This represents a shift toward modular, data-generating forms of instruction that can be more easily monitored and analyzed. All the aforementioned tools track and record minute student data, including their keystrokes, time spent on a quiz, or movement during an exam.[12] Several universities, including Texas A&M[13] and the University of Arizona,[14] are even

7. PAM ARROWAY, GLENDA MORGAN, MOLLY O'KEEFE & RONALD YANOSKY, LEARNING ANALYTICS IN HIGHER EDUCATION 10 (2016), https://library.educause.edu/~/media/files/library/2016/2/ers1504la.

8. *See* Neil Selwyn, *What's the Problem with Learning Analytics?*, 6 J. LEARNING ANALYTICS 11, 13–14 (2019).

9. *LockDown Browser*, RESPONDUS, https://web.respondus.com/he/lockdownbrowser (last visited May 30, 2025).

10. PROCTORU, https://www.proctoru.com (last visited May 30, 2025).

11. *See Policy for Use of a Learning Management System for Online Classes*, CUNY, https://commons.hostos.cuny.edu/edtech/policy/policy-for-use-of-a-learning-management-system-for-online-classes/(last visited Aug. 10, 2025).

12. Zak Vescera, *Canvas Is Tracking Your Data: What Is UBC Doing with It?*, UBYSSEY (Mar. 27, 2019), https://ubyssey.ca/features/double-edged-sword/; Shanay Murdock, *How Do I Track Student Activity in My Course?*, FSU (Oct. 23, 2024), https://support.canvas.fsu.edu/kb/article/893-how-do-i-track-student-activity-in-my-course/; *see* E.A. Kochegurova & R.P. Zateev, *Hidden Monitoring Based on Keystroke Dynamics in Online Examination System*, 48 PROGRAMMING & COMPUT. SOFTWARE 385, 386 (2022); Mario Garcia Valdez, Juan-J. Merelo, Amaury Hernandez Aguila & Alejandra Mancilla Soto, *Mining of Keystroke and Mouse Dynamics to Increase the Engagement of Students with Programming Assignments*, 829 STUD. COMPUTATIONAL INTEL. 41, 57–58 (2019); Haotian Li, Min Xu, Yong Wang, Huan Wei & Huamin Qu, *A Visual Analytics Approach to Facilitate the Proctoring of Online Exams*, CHI '21: PROC. OF THE 2021 CHI CONF. ON HUM. FACTORS COMPUTING SYS. 1, 15 (May 2021).

13. Hannah Conrad, *Texas A&M Researchers Partner with MoodMe to Enhance Facial Analysis*, TEX. A&M UNIV. ENG'G (Aug. 5, 2019), https://engineering.tamu.edu/news/2019/08/texas-am-researchers-partner-with-moodme-to-enhance-facial-analysis.html.

14. *Using Facial Recognition Software to Measure Emotions*, UNIV. ARIZ. ELLER COLL. MGMT. (Jan. 3, 2020), https://eller.arizona.edu/news/2020/01/using-facial-recognition-software-measure-emotions.

researching the use of AI for emotion recognition to assess students' attentiveness and emotional responses during instruction.

The implementation of technology-based instruction has allowed pedagogical decisions to be increasingly influenced by institutional priorities, such as course completion rates, student satisfaction rates, and the cost per student. The emergence of generative AI tools has further complicated the role of technology in education. Instructors must now contend with the lure of AI detection software as well as AI-assisted and fully automated grading tools, despite ethical and pedagogical concerns about generative AI hallucinations and AI's inability to appropriately evaluate context, creativity, and nuance.[15] Instructors are also tasked with "AI-proofing" assignments and assessments, leading to professor- and course-specific AI-use policies which create a fragmented learning environment for students who must navigate varying AI policies from course to course.[16] The pedagogical implications of generative AI also remain under-researched, but emerging scholarship suggests that many AI lesson plan generators embed outdated educational methods which limit academic dialogue and student freedom.[17] The rapid adoption of such novel and under-examined technology across classrooms reflects an impulse of higher education institutions that more data, more automation, and more rapid adoption of emerging, unexamined technologies inherently improve student outcomes and learning experiences.

Beyond classroom instruction, technology is integrated into university operations and thus student life. Campus infrastructure increasingly relies on network systems that automate access and facility management. Students routinely use identification cards or mobile credentials to access dormitories, academic buildings, dining halls, recreation facilities, and libraries. Many schools have integrated technology into dormitories, such as Amazon Echo Dots and Wi-Fi,[18] by default. Traffic through university Wi-Fi is not private

15. *See* Yuhan Gao, *AI and Auto-Grading in Higher Education: Capabilities, Ethics, and the Evolving Role of Educators*, OHIO ST. ASC OFF. OF DISTANCE EDUC. (July 15, 2025), https://ascode.osu.edu/news/ai-and-auto-grading-higher-education-capabilities-ethics-and-evolving-role-educators.

16. *See* Christopher Rim, *How Ivy League Schools Are Navigating AI in the Classroom*, FORBES (July 7, 2025), https://www.forbes.com/sites/christopherrim/2025/07/07/how-ivy-league-schools-are-navigating-ai-in-the-classroom/.

17. *See* Bodong Chen, Jiayu Cheng, Chen Wang & Vivian Leung, *Pedagogical Biases in AI-Powered Educational Tools: The Case of Lesson Plan Generators*, 30 SOC. INNOVATIONS J. 2–3 (2025).

18. *See Amazon Echos to Be Installed in Dorms*, UTD MERCURY (Apr. 15, 2019), https://utdmercury.com/amazon-echos-to-be-installed-in-dorms/; Molly Price, *Alexa, Time for Class: How One University Put an Echo Dot in Every Dorm Room*, CNET (Aug. 13, 2019), https://www.cnet.com/home/smart-home/features/alexa-time-for-class-how-one-university-put-an-echo-dot-in-every-dorm-room/; Reisman, *supra* note 1, at 565.

and can be monitored.[19] Each swipe, tap, device use, or signal gives universities and third-party vendors access to students' locational data, habits, behaviors, and activities on a day-to-day basis.[20] Some even track dining hall entry and time spent eating as a measure of social connectedness, flagging students for intervention if they appear isolated.[21] Technologies are also used in campus security, such as surveillance cameras, emergency alert systems, access control tools, and facial recognition software. Several universities, including Columbus State University, Florida International University, Iowa State University, the University of Alabama, the University of Illinois, and the University of Wisconsin, utilize facial recognition technology.[22] These operational systems, although aimed at enhancing convenience or safety, contribute to the everyday monitoring of students.

Research suggests that students generally express a high level of trust in their institutions, particularly nonprofit institutions.[23] Students distinguish between data collection in commercial contexts versus educational ones, assuming universities operate under a duty of care.[24] However, students also view their relationships with their universities as transactional. Students accept their disclosure of data as part of their relationship with their university.[25] This

19. David Galef, *The New Hall Monitor*, INSIDE HIGHER ED (July 10, 2024), https://www.insidehighered.com/opinion/views/2024/07/10/when-you-use-campus-wi-fi-whos-watching-and-how-opinion.

20. WEINBERG, *supra* note 4, at 69–70, 73, 77; Reisman, *supra* note 1, at 563–65.

21. Vimal Patel, *Are Students Socially Connected? Check Their Dining-Hall-Swipe Data*, CHRON. HIGHER EDUC. (Apr. 9, 2019), https://www.chronicle.com/article/are-students-socially-connected-check-their-dining-hall-swipe-data/; Nicholas A. Bowman, Lindsay Jarratt, Linnea A. Polgreen, Thomas Kruckeberg & Alberto M. Segre, *Early Identification of Students' Social Networks: Predicting College Retention and Graduation via Campus Dining*, 60 J. COLL. STUDENT DEV. 617, 620–22 (2019); *see* WEINBERG, *supra* note 4, at 73.

22. *Stop Facial Recognition on Campus*, FIGHT FOR THE FUTURE, campus.banfacialrecognition.com/ (last visited Aug. 12, 2025).

23. *See, e.g.*, Kyle M.L. Jones, Andrew Asher, Abigail Goben, Michael R. Perry, Dorothea Salo, Kristin A. Briney & M. Brooke Robertshaw, *"We're Being Tracked at All Times": Student Perspectives of Their Privacy in Relation to Learning Analytics in Higher Education*, 71 J. ASS'N INFO. SCI. & TECH. 1044, 1053–54 (2020); Jasmine Park & Amelia Vance, *Data Privacy in Higher Education: Yes, Students Care*, EDUCAUSE REV. (Feb. 11, 2021), https://er.educause.edu/articles/2021/2/data-privacy-in-higher-education-yes-students-care; Melissa Ezarik, *Data Collection Comforts: Most Students Trust Their Colleges*, INSIDE HIGHER ED (Aug. 16, 2021), https://www.insidehighered.com/news/2021/08/17/nine-ways-raise-awareness-about-student-data-and-data-privacy.

24. Kyle M.L. Jones, Alan Rubel & Ellen LeClere, *A Matter of Trust: Higher Education Institutions as Information Fiduciaries in an Age of Educational Data Mining and Learning Analytics*, 71 J. ASS'N INFO. SCI. & TECH. 1227, 1230 (2019).

25. *See* Kyle M.L. Jones, Abigail Goben, Michael R. Perry, Mariana Regalado, Dorothea Salo, Andrew D. Asher, Maura A. Smale & Kristin A. Briney, *Transparency and Consent: Student Perspectives on Educational Data Analytics Scenarios*, 23 LIBRS. & ACAD. 485, 499 (2023); Sharon

perceived lack of choice may be representative of the pressure that universities place on students to use campus technology. As more campus infrastructure is controlled through digital technology, students have fewer recourses to choose not to use digital technology. For example, students may not be able to use prepaid dining dollars on meal plans without downloading the GrubHub app. Yet doing so exposes students to intensive data collection: GrubHub collects and shares geolocational and behavioral data to deliver targeted advertisements.[26] Opting out of such on-campus technologies can mean forgoing basic services, while opting in subjects students to invasive tracking.

Universities do not always provide information to students about how technology is used or what data is collected and shared. At the University of North Carolina, attendance tracking sensors from SpotterEDU were installed without notification to faculty or students, causing mass confusion and leading to one dean physically removing a sensor.[27] In many cases, higher education institutions have failed to promote informed consent practices within and outside of classrooms.[28] For example, in 2022, George Washington University apologized to students, staff, and faculty for monitoring their locations without obtaining consent.[29]

Universities have also used technologies to supplement in-person services offered on campus. Many colleges offer mental health apps such as Welltrack to students for free.[30] The app prompts students to log emotions and upsetting experiences through tools like the "Thought Diary."[31] This data is then aggregated and presented in real-time dashboards for university

Slade, Paul Prinsloo & Mohammad Khalil, *Learning Analytics at the Intersections of Student Trust, Disclosure and Benefit*, 2019 9TH INT'L LEARNING ANALYTICS & KNOWLEDGE CONF. 235, 240.

26. *Privacy Policy*, GRUBHUB, https://www.grubhub.com/legal (last visited July 8, 2025).

27. Wesley Jenkins, *Some Colleges Are Using Beacon Technology to Track Athletes' Attendance. Is That Ethical?*, CHRON. HIGHER EDUC. (Nov. 7, 2019), https://www.chronicle.com/article/some-colleges-are-using-beacon-technology-to-track-athletes-attendance-is-that-ethical/; Drew Harwell, *Colleges Are Turning Students' Phones into Surveillance Machines, Tracking the Locations of Hundreds of Thousands*, WASH. POST (Dec. 24, 2019), http://www.washingtonpost.com/technology/2019/12/24/colleges-are-turning-students-phones-into-surveillance-machines-tracking-locations-hundreds-thousands/.

28. Kyle M.L. Jones, *Learning Analytics and Higher Education: A Proposed Model for Establishing Informed Consent Mechanisms to Promote Student Privacy and Autonomy*, 16 INT'L J. EDUC. TECH. HIGHER EDUC. 8–9 (2019); R.J. Connelly, *Intentional Learning: The Need for Explicit Informed Consent in Higher Education*, 49 J. GEN. EDUC. 211, 225–29 (2000).

29. Angela Brown, *University Apologizes for Data Project That Monitored Students Location*, NAT'L DESK (Feb. 21, 2022), https://thenationaldesk.com/news/americas-news-now/university-apologizes-for-not-informing-campus-wide-community-it-was-collecting-their-data.

30. WEINBERG, *supra* note 4, at 83–84, 91; *see Give Your Campus a Boost*, WELLTRACK BOOST, https://welltrack-boost.com/higher-education/ (last visited May 30, 2025).

31. WELLTRACK BOOST, https://welltrack-boost.com (last visited May 30, 2025).

administrators.[32] A Welltrack sales representative, alongside promotional materials, describes the app as a solution to the problem of the "overutilization" of campus mental health resources.[33] In this way, technology implementation on campus may be driven by resource constraints instead of concern for students. For example, San Jose State University's (SJSU) 2013 partnership with the online education company Udacity was framed as a way to reduce costs, claiming to offer students "*college classes for credit from an accredited university at a very affordable price* of $150 per course."[34] SJSU and Udacity entered into a revenue-sharing agreement, reflecting a financial incentive to scale up online instruction.[35] SJSU ultimately suspended the program due to disappointing outcomes: pass rates ranged from just 20 to 44 percent, compared to around 75 percent in the equivalent face-to-face classes.[36]

Students often lack the technology, data, and privacy literacy needed to critically assess or assert preferences about digital technologies on campus.[37] Structural barriers to transparency, such as opacity about what technologies were procured and deployed, complex privacy policies,[38] and limited access to information from universities make it difficult for students to understand what technologies are in use and how their data is collected or used. Despite informational gaps, students often do express privacy concerns that align with established privacy principles, including concerns about data granularity, access, and secondary use.[39] Generally, many Americans lack the knowledge and autonomy required to meaningfully consent to data collection, casting doubt on current consent frameworks.[40] Still, students do act on their concerns: at Virginia Commonwealth University, when students learned about a new attendance tool that analyzes student Wi-Fi connectivity, over 50 percent of students opted out within the two-month window.[41]

32. WEINBERG, *supra* note 4, at 108.

33. *Id.*; *Welltrack Connect for Universities*, WELLTRACK CONNECT, https://welltrack-connect.com/universities (last visited May 30, 2025).

34. Pat Lopes Harris, *SJSU and Udacity Partnership*, SJSU NEWSCENTER (Jan. 15, 2013), https://blogs.sjsu.edu/newsroom/2013/sjsu-and-udacity-partnership/.

35. *Id.*

36. Gregory Ferenstein, *San Jose's Bold Experiment in Online Ed Disappoints, Suspends Pilot with Udacity*, TECHCRUNCH (July 19, 2013), https://techcrunch.com/2013/07/19/san-jose-states-bold-experiment-in-online-ed-disappoints-suspends-pilot-with-udacity/.

37. Park & Vance, *supra* note 23.

38. *Id.*

39. Jones et al., *supra* note 23, at 1051–53.

40. JOSEPH TUROW, YPHTACH LELKES, NORA A. DRAPER & ARI EZRA WALDMAN, AMERICANS CAN'T CONSENT TO COMPANIES' USE OF THEIR DATA 8–14 (2023), https://www.asc.upenn.edu/sites/default/files/2023-02/Americans_Can%27t_Consent.pdf.

41. Reisman, *supra* note 1, at 590–91.

Students also bear the burden of discrimination or surveillance perpetuated by digital technologies implemented at universities.[42] At Temple University, an early-alert system was designed in part by a former criminologist who previously worked on recidivism prediction tools, raising concerns about the application of carceral logic to educational settings.[43] Similarly, proctoring services are known to flag people of color and people with disabilities at a higher rate, singling these individuals out for punitive measures.[44] These technologies reflect broader demonstrated racial bias in facial and emotion recognition technologies, many of which exhibit higher error rates when applied to non-White subjects.[45]

Moreover, increased surveillance can also influence students' behavior and sense of autonomy. Students reported modifying their use of campus resources in response to monitoring technologies, such as self-censoring their searches on campus Wi-Fi.[46] Surveillance during assessments through proctoring software can lead students to alter their test-taking behaviors out of fear of triggering automated flags, which worsens testing conditions and causes significant anxiety.[47] Similarly, the presence of facial recognition or emotion recognition systems in classrooms and public spaces may cause students to change their behavior and activities.[48] These effects are not evenly distributed:

---

42. Chris Gilliard & Neil Selwyn, *Automated Surveillance in Education*, 5 POSTDIGITAL SCI. & EDUC. 195, 201–02 (2023).

43. *See* EKOWO & PALMER, *supra* note 6, at 8.

44. Simon Coghlan, Tim Miller & Jeannie Paterson, *Good Proctor or "Big Brother"? Ethics of Online Exam Supervision Technologies*, 34 PHIL. & TECH. 1581, 1591–92 (2021); Lindsay McKenzie, *Proctoring Tool Failed to Recognize Dark Skin, Students Say*, INSIDE HIGHER ED (Apr. 5, 2021), https://www.insidehighered.com/quicktakes/2021/04/06/proctoring-tool-failed-recognize-dark-skin-students-say; Deborah R. Yoder-Himes, Alina Asif, Kaelin Kinney, Tiffany J. Brandt, Rhiannon E. Cecil, Paul R. Himes, Cara Cashon, Rachel M.P. Hopp & Edna Ross, *Racial, Skin Tone, and Sex Disparities in Automated Proctoring Software*, 7 FRONTIERS EDUC. 4–8 (2022); Shea Swauger, *Software That Monitors Students During Tests Perpetuates Inequality and Violates Their Privacy*, MIT TECH. REV. (Aug. 7, 2020), https://www.technologyreview.com/2020/08/07/1006132/software-algorithms-proctoring-online-tests-ai-ethics/; Katie Ignatowski, *Surveillance Tech Is Wrongly Accusing Disabled Students of Cheating on Tests*, TRUTHOUT (June 9, 2022), https://truthout.org/articles/surveillance-tech-is-wrongly-accusing-disabled-students-of-cheating-on-tests/.

45. Michael Kwet & Paul Prinsloo, *The 'Smart' Classroom: A New Frontier in the Age of the Smart University*, 25 TEACHING HIGHER EDUC. 510, 519 (2020).

46. Jones et al., *supra* note 23, at 1052.

47. Annika Pokorny, Cissy J. Ballen, Abby Grace Drake, Emily P. Driessen, Sheritta Fagbodun, Brian Gibbens, Jeremiah A. Henning, Sophie J. McCoy, Seth K. Thompson, Charles G. Willis & A. Kelly Lane, *"Out of My Control": Science Undergraduates Report Mental Health Concerns and Inconsistent Conditions When Using Remote Proctoring Software*, 19 INT'L J. EDUC. INTEGRITY, Nov. 15, 2023, at 17.

48. Hengyi Fu & Yao Lyu, *Facial Recognition Interaction in a University Setting: Impression, Reaction, and Decision-Making*, 13192 LECTURE NOTES COMPUT. SCI. 329, 333–35 (2022).

students from historically marginalized groups are more likely to experience surveillance as a source of discomfort or exclusion, particularly when the technologies in question have documented patterns of racial or linguistic bias.[49]

The educational consequences for students are also notable. Instructors and advisors may use predictive analytics to steer students toward courses in which they are statistically more likely to succeed. Even more extreme, if systems predict that a student is unlikely to be retained, administrators and advisors may lessen their allocation of time and resources to that student or encourage them to withdraw entirely.[50] This is not entirely unrealistic: the president of Mount St. Mary's University used predictive analytics to identify struggling first-year students and offered tuition refunds for those choosing to leave before the cutoff date for reporting the school's enrollment to the federal government so that the school's retention rate would not suffer.[51] Defending the policy and use of technology, the president wrote to faculty, "This is hard for you because you think of the students as cuddly bunnies, but you can't. You just have to drown the bunnies . . . put a Glock to their heads."[52] While such predictive practices improve retention and course completion metrics, they may also limit students' exposure to intellectually challenging or professionally relevant coursework. These systems encourage a predictable and efficient course of completion for a college degree that leaves little room for intellectual experimentation or exploration.

Digital technologies are integrated into universities based on the promise of increased administrative efficiency, which is not always aligned with improved learning outcomes or positive student experiences.[53] Additionally, university rankings tend to measure institutional resources and selectivity rather than learning outcomes, instruction, and affordability.[54] The data that

49. *See* Mona Sloane, *Surveillance Society: Artificial Lighting for a Policed Public*, ARCHITECTURAL REV. (Sep. 15, 2021), https://www.architectural-review.com/essays/technology/surveillance-society-artificial-lighting-for-a-policed-public; Allison Koenecke, Andrew Nam, Emily Lake, Joe Nudell, Minnie Quartey, Zion Mengesha, Connor Toups, John R. Rickford, Dan Jurafsky & Sharad Goel, *Racial Disparities in Automated Speech Recognition*, 117 PNAS 7684, 7687 (2020) (concluding that "it is considerably harder for African Americans to benefit from . . . speech recognition technology").

50. Kyle M.L. Jones & Chase McCoy, *Reconsidering Data in Learning Analytics: Opportunities for Critical Research Using a Documentation Studies Framework*, 44 LEARNING MEDIA & TECH. 52, 56 (2019).

51. *Id.*; Scott Jaschik, *Are At-Risk Students Bunnies to Be Drowned?*, INSIDE HIGHER ED (Jan. 19, 2016), https://www.insidehighered.com/news/2016/01/20/furor-mount-st-marys-over-presidents-alleged-plan-cull-students.

52. Jaschik, *supra* note 51.

53. Zoia Sharlovych, Liudmyla Vilchynska, Serhiy Danylyuk, Bohdan Huba & Halyna Zadilska, *Digital Technologies as a Means of Improving the Efficiency of Higher Education*, 13 INT'L J. INFO. & EDUC. TECH. 1214, 1215–19 (2023).

54. ROBERT KELCHEN, HIGHER EDUCATION ACCOUNTABILITY (2018).

students produce as part of participating in university life is often used as business intelligence and as a means of evaluating institutional effectiveness and funding.[55]

This means that students now leave behind extensive digital footprints that may influence future risk assessments, employment opportunities, visa eligibility, or insurance premiums as student data is sold by vendors to analytics firms outside of higher education, for example as training data for AI models. Many technology providers, including the LMS platform Canvas, are criticized for their vague and opaque privacy policies.[56] Though companies often assert that they do not sell student data, acquisitions and mergers complicate these assurances.[57] For example, the private equity firm Thoma Bravo acquired Instructure, the company behind Canvas, followed by another acquisition by KKR.[58] Instructure is now the subject of a class-action lawsuit alleging violations of the privacy rights of minor students, with claims that the platform monetized student data and shared it with third parties.[59]

These developments point to a broader structural issue: universities have ceded substantial control over student data and digital technologies to an ecosystem of private vendors that operate with little transparency. Technology vendors disclose only minimal, opaque information as to how they use the data they collect from users.[60] Many services have further data outflow to third parties, whether through the sharing of advertising preferences or students' personally identifiable information. For example, Pearson's MyLab, a digital learning platform, transmits student names and emails to Google Analytics, along with notifications of what students read and highlight in their digital

55. Ben Williamson, *The Hidden Architecture of Higher Education: Building a Big Data Infrastructure for the 'Smarter University,'* 15 INT'L J. EDUC. TECH. HIGHER EDUC., Mar. 8, 2018, at 10.

56. *See, e.g.*, Britt Paris, Rebecca Reynolds & Catherine McGowan, *Platforms Like Canvas Play Fast and Loose with Students' Data*, NATION (Apr. 22, 2021), https://www.thenation.com/article/society/canvas-surveillance/.

57. *See* Matthew Rozsa, *Students Fear for Their Data Privacy After University of California Invests in Private Equity Firm*, SALON (July 28, 2020), https://www.salon.com/2020/07/28/students-fear-for-their-data-privacy-after-university-of-california-invests-in-private-equity-firm/.

58. Press Release, Instructure, Instructure to be Acquired by KKR (July 25, 2024), https://www.instructure.com/press-release/instructure-to-be-acquired-by-KKR.

59. Class Action Complaint, Hernandez-Silva v. Instructure, Inc., No. 25-cv-02711 (C.D. Cal. Mar. 27, 2025), ECF No. 1; *see* Roma Patel, *EdTech and Privacy of Student Information: A Case Study*, ROBINSON+COLE (Apr. 3, 2025), https://www.dataprivacyandsecurityinsider.com/2025/04/edtech-and-privacy-of-student-information-a-case-study/.

60. Michele Molnar, *Most Ed-Tech Products Don't Meet Minimum Criteria in Their Privacy Policies, Report Finds*, EDWEEK MKT. BRIEF (May 29, 2018), https://marketbrief.edweek.org/regulation-policy/most-ed-tech-products-dont-meet-minimum-criteria-in-their-privacy-policies-report-finds/2018/05.

textbook.[61] These third-party platforms do not have agreements with the university itself and oftentimes the university is not aware of this data outflow.

As a result of the COVID-19 pandemic, universities became increasingly reliant on digital technologies in every aspect, from contact tracing to lecture delivery to remote food ordering.[62] Commercial technology solutions were rapidly procured en masse, with a decreased opportunity to review these tools or for university stakeholders to resist their adoption.[63] This expansion has become normalized and entrenched as many of these digital technologies are widely used at universities today.[64] As universities outsource the provision of services to vendors, technologies introduced as supplemental tools often evolve into infrastructural necessities.

Once a tool or platform is adopted, it is rarely discontinued even if its utility remains unproven. Instead, vendors can introduce new features and integrations to expand their data collection without sufficient institutional oversight. Most universities lack the capacity to continuously monitor and assess their hundreds of procured digital technologies. When SpotterEDU was procured at Syracuse University, it contained a feature allowing students to share their exact GPS coordinates with faculty. While this feature was later removed, it was part of the original license despite its irrelevance to SpotterEDU's intended purpose of class attendance tracking.[65] The unchecked expansion of features and tools from vendors allows them to embed commercial priorities into the core of the university over other stakeholders and their concerns. Vendors have different priorities, such as mitigating technical deficits rather than addressing the broader technology concerns held

61. Taylor Swaak, *The 'Textbook' That Reads You*, CHRON. HIGHER EDUC. (July 20, 2023), https://www.chronicle.com/article/the-textbook-that-reads-you.

62. Kristin R.V. Harrington, Meron R. Siira, Elizabeth P. Rothschild, Sharon R. Rabinovitz, Samuel Shartar, David Clark, Alexander Isakov, Allison Chamberlain, Enku Gelaye, J. Peter Cegielski & Neel R. Gandhi, *A University-Led Contact Tracing Program Response to a COVID-19 Outbreak Among Students in Georgia, February-March 2021*, 137 PUB. HEALTH REPS. 61S (2022); Darrell J.R. Evans, *Has Pedagogy, Technology, and Covid-19 Killed the Face-to-Face Lecture?*, 15 ANATOMICAL SCIS. EDUC. 1145 (2022); Press Release, Grubhub, Grubhub and Transact Partner to Offer Universities Expanded Off-Campus Meal Spending Programs for Students (Aug. 12, 2021), https://about.grubhub.com/news/grubhub-and-transact-partner-to-offer-universities-expanded-off-campus-meal-spending-programs-for-students/.

63. Kyle M.L. Jones, Amy VanScoy, Alison Harding & Amy Martin, *Changing Student Privacy Responsibilities and Governance Needs: Views from Faculty, Instructional Designers, and Academic Librarians*, 37 J. COMPUTING HIGHER EDUC. 327, 340 (2025); Janja Komljenovic, *The Future of Value in Digitalised Higher Education: Why Data Privacy Should Not Be Our Biggest Concern*, 83 HIGHER EDUC. 119, 120 (2022).

64. Nurullah Aydın, Muhammed Fatih Sayır, Süleyman Aydeniz & Tacettin Şimşek, *How Did COVID-19 Change Faculty Members' Use of Technology?*, 13 SAGE OPEN, Jan. 18, 2023, at 8–9.

65. Harwell, *supra* note 27.

by faculty.[66] However, assuming, arguendo, that a university has full awareness of vendors' uses of data, existing governance structures give students and faculty little control. We discuss such governance structures in the following Part.

III. TECHNOLOGY GOVERNANCE IN UNIVERSITIES

Technology governance at universities has aligned with developments in university governance more generally. In many cases, administrators dictate most policies with minimal faculty input and rarely any student involvement.

In higher education, technology governance encompasses both IT and data concerns and focuses on the structures, processes, and policies that direct the procurement, use, and oversight of digital systems and data in support of administrative, teaching, and research activities.[67] Experts suggest that effective technology governance practices, including ensuring information security, managing digital integrations, and protecting privacy, must match the organization's strategic mission.[68] In this context, effectiveness refers to how well governance mechanisms ensure that technology is used to deliver institutional value, manage risks, and optimize resources in alignment with the university's broader mission.[69]

In most universities, governance responsibilities are distributed across a hierarchy of stakeholders from board-level positions to departmental units. At the highest level, board-level IT or technology committees set the strategic direction of university technology governance. These groups play a key role in ensuring that technology mirrors the institution's objectives. Recently, university executive leadership has also expanded to include positions such as

66. *See* Emma Harvey, Allison Koenecke & Rene F. Kizilcec, *"Don't Forget the Teachers": Towards an Educator-Centered Understanding of Harms from Large Language Models in Education*, CHI '25: PROC. OF THE 2025 CHI CONF. ON HUM. FACTORS COMPUTING SYS. 1, 9 (Apr.–May 2025).

67. Michael Hicks, Graham Pervan & Brian Perrin, *A Case Study of Improving Information Technology Governance in a University Context*, 318 IFIP ADVANCES INFO. & COMMC'N TECH. 89, 89–90 (2010); Alejandra Oñate-Andino, David Mauricio, Gloria Arcos-Medina & Danilo Pastor, *The Application and Use of Information Technology Governance at the University Level*, 858 ADVANCES INTELLIGENT SYS. & COMPUTING 1028, 1028–29 (2018).

68. *See, e.g.*, Peter Weill & Jeanne W. Ross, *IT Governance on One Page* 6–7 (MIT Ctr. for Info. Sys. Rsch., Working Paper No. 349, 2004); PETER WEILL & JEANNE W. ROSS, IT GOVERNANCE: HOW TOP PERFORMERS MANAGE IT DECISION RIGHTS FOR SUPERIOR RESULTS 2–3 (2004); Deborah Louise Carraway, Information Technology Governance Maturity and Technology Innovation in Higher Education: Factors in Effectiveness 17 (May 1, 2015) (M.S. thesis, University of North Carolina at Greensboro), https://libres.uncg.edu/ir/uncg/listing.aspx?id=18073.

69. Carraway, *supra* note 68, at 29–30.

Chief Information Officer (CIO), Chief Information Security Officer (CISO), and Chief Privacy Officer (CPO).[70]

Universities additionally rely on standing committees, councils, or working groups to coordinate technology governance across departments and administrative units, especially to establish data classification standards, oversee compliance with privacy and security regulations, and develop policy recommendations. Approximately 75 percent of universities have a data governance body in this form.[71] Many have also established campus-wide privacy governance boards, including UCLA's Board on Privacy and Data Protection and the University of Chicago's Data Stewardship Council.[72]

At the management and implementation levels, IT offices are central to the execution of technology governance practices. On average, IT offices are responsible for around 60 percent of technology functions and services across universities.[73] As institutions move toward centralized technology resource management, IT offices perform the critical functions of procuring technology, implementing cybersecurity protocols, and providing guidance to faculty, students, and administrative units.[74]

Institutional research (IR) offices or attachments play a similar role in managing and operationalizing data-based resources. Typically the second- or third-largest team involved in data governance,[75] IR teams collect, analyze, and steward institutional data. Often embedded within business, planning, enrollment and admissions, or financial units,[76] institutional researchers facilitate data analytics operations, including predictive analytics and learning analytics, for other teams at universities. IR offices are frequently tasked with ensuring that data privacy and protection standards are being followed.

70. Merritt Neale & Matthew Tryniecki, *The Post-Pandemic Evolution of Student Data Privacy*, EDUCAUSE REV. 51, 55–56 (2020), https://er.educause.edu/-/media/files/articles/2020/8/er20_3104.pdf; Mike Wulff, *The Evolving Role of CIOs in Higher Education*, EDUCAUSE REV., Aug. 25, 2022, https://er.educause.edu/articles/2022/8/the-evolving-role-of-cios-in-higher-education; Valerie Vogt, *The Chief Privacy Officer in Higher Education,* EDUCAUSE REV., June 4, 2018, https://er.educause.edu/articles/2018/6/its-time-to-set-cisos-free.

71. Cary K. Jim & Hsia-Ching Chang, *The Current State of Data Governance in Higher Education*, 55 PROC. ASS'N INFO. SCI. & TECH. 198, 202 (2018).

72. Neale & Tryniecki, *supra* note 70, at 56.

73. Che-Wei Liu, Peng Huang & Henry C. Lucas Jr., *Centralized IT Decision Making and Cybersecurity Breaches: Evidence from U.S. Higher Education Institutions*, 37 J. MGMT. INFO. SYS. 758, 772 (2020).

74. *What Is the Higher Education IT Environment?*, UNIV. WASH., https://uwconnect.uw.edu/it?id=kb_article_view&sysparm_article=KB0034193 (last visited May 30, 2025).

75. Jim & Chang, *supra* note 71, at 201.

76. J. Fredericks Volkwein, *The Foundations and Evolution of Institutional Research*, 2008 NEW DIRECTIONS HIGHER EDUC. 5, 6.

Enforcement of the Family Educational Rights and Privacy Act (FERPA) often falls to IR offices.[77]

Even though IR offices are supposed to enforce existing regulations like FERPA, these regulations have failed to keep pace with the development of the digital technology ecosystem in higher education. FERPA guarantees students the right to inspect their education records, amend inaccurate or misleading information, and file complaints if their rights are violated.[78] However, institutions are not equipped to track information flow to dozens or hundreds of technology vendors, much less inform students of how data about them is used, to what ends, and by whom.[79] FERPA also allows universities to disclose private identifiable information about students without their consent or knowledge to any third party that provides "institutional services or functions," like an educational technology company.[80] As long as a "legitimate educational interest" exists, the sharing and use of student data is allowed.[81] Universities have broad discretion in determining who qualifies as having a "legitimate educational interest" in student data.[82] FERPA provides no clear standards or guidance for what constitutes such an interest, effectively allowing institutions to classify nearly any third-party vendor as eligible to receive student information.[83]

FERPA's original intent was to prevent inaccurate information from permanently following students.[84] It now enables the opposite: data flows out of the university to private vendors with minimal oversight or restriction. In practice, FERPA delegates most decision-making and enforcement regarding student privacy to educational institutions, which in turn defer to private vendors. In doing so, universities place student data in the hands of those with a commercial interest in sharing and selling that data.[85]

---

77. Kyle M.L. Jones & Chase McCoy, *Privacy in Practice: A Socio-Technical Integration Research (STIR) Study of Rules-in-Use Within Institutional Research*, *in* GOVERNING PRIVACY IN KNOWLEDGE COMMONS 98, 107 (Madelyn Rose Sanfilippo, Brett M. Frischmann & Katherine J. Strandburg eds., 2021).

78. Family Educational Rights and Privacy Act of 1974, 20 U.S.C. § 1232g(a)(1)–(2), (f)–(g).

79. Elana Zeide, *The Limits of Education Purpose Limitations*, 71 U. MIAMI L. REV. 494, 503–09 (2017).

80. 34 C.F.R. § 99.31(a)(1)(i)(B) (2025); Jones, *supra* note 28, at 10.

81. Jones et al., *supra* note 25, at 502.

82. *Id.*

83. Michael Brown & Carrie Klein, *Whose Data? Which Rights? Whose Power? A Policy Discourse Analysis of Student Privacy Policy Documents*, 91 J. HIGHER EDUC. 1149, 1165 (2020).

84. Zeide, *supra* note 79, at 499, 501, 503, 514.

85. Elana Zeide, *Student Privacy Principles for the Age of Big Data: Moving Beyond Ferpa and Fipps*, 8 DREXEL L. REV. 339, 342, 359–62 (2016).

Moreover, formal training on legal requirements for privacy including FERPA remains inconsistent among institutional researchers. According to a survey of 232 institutional researchers, 53 percent taught themselves about FERPA, while 22.5 percent received no training.[86] Fuller also highlights a culture in institutional research that prioritizes individual practice over formal written policies, resulting in inconsistent FERPA enforcement.[87] This lack of training and documentation exposes students to greater harm in the event of a data breach or FERPA violation.

Although not always viewed as central actors in technology governance, faculty exert influence over the adoption and use of learning technologies. Requests to IT offices to procure or allow new educational tools often originate with faculty.[88] However, faculty do not consider themselves to be tasked with upholding student privacy. While many faculty believe privacy is important to intellectual freedom and the learning environment, this concern does not translate into the selection of privacy-protecting technologies. Instead, faculty tend to rely on institutional vetting offices, trusting that the tools approved by the university meet the necessary legal and ethical requirements.[89]

University legal counsel also plays a pivotal yet underexamined role in technology governance. Legal offices are responsible for drafting, reviewing, and approving procurement contracts with third-party service providers, including data sharing and user agreements. They coordinate institutional efforts to ensure regulatory compliance with relevant privacy laws.[90] However, the designation of legal counsel as arbiters of technology decisions may lead to a governance framework that emphasizes compliance and liability minimization over ethical considerations such as informed consent.[91]

Universities are a specialized context for IT governance, which grew out of the corporate context.[92] Existing frameworks for technology governance, such as Information Technology Infrastructure Library (ITIL)[93] or Control

86. Jones & McCoy, *supra* note 77, at 107.

87. Matthew Fuller, *The Practices, Policies, and Legal Boundaries Framework in Assessment and Institutional Research*, 2016 NEW DIRECTIONS FOR INSTITUTIONAL RSCH. 9, 23 (2017).

88. Swaak, *supra* note 61.

89. Kyle M.L. Jones, Amy VanScoy, Kawanna Bright, Alison Harding & Sanika Vedak, *A Measurement of Faculty Views on the Meaning and Value of Student Privacy*, 34 J. COMPUTING HIGHER EDUC. 769, 782–83 (2022).

90. David Jesse, *Your College's Top Lawyer Has Never Been More Powerful*, CHRON. HIGHER EDUC. (Feb. 26, 2024), https://www.chronicle.com/article/your-colleges-top-lawyer-has-never-been-more-powerful.

91. Jones et al., *supra* note 63, at 341.

92. Oñate-Andino et al., *supra* note 67, at 1029.

93. ITIL is a framework that provides standardized procedures and best practices for managing IT services, originally developed for use in commercial and government sectors.

Objectives for Information and Related Technologies (COBIT),[94] are frequently ill-suited to at least some of the needs of academic institutions.[95] Generally, literature suggests that none of the prominent technology frameworks comprehensively cover process, structure, human behavior, and organizational culture.[96] Accordingly, many universities create their own frameworks for governance, which bring other unique challenges. There is a contemporary lack of research on information technology governance in universities, which complicates efforts to create robust frameworks.[97] This dearth of research is also true for the technology governance of big data and AI in universities, which, as discussed, are increasingly ubiquitous.[98]

The decentralization of many universities also poses a challenge to cohesive technology governance. CIOs and CISOs point to the culture of decentralization as one of the major barriers to information security.[99] This also helps contextualize the importance of committees, councils, and working groups within universities to bring together decision-makers across administrative units. Universities with a higher degree of centralization suffer fewer cybersecurity breaches, and this effect is strongest in public and research universities.[100]

Many universities also suffer from insufficient training among governance stakeholders. Board members rarely come from technical backgrounds, so they lack the expertise needed to understand the implications of technology

---

Sarah K. White & Lynn Greiner, *What Is ITIL? Your Guide to the IT Infrastructure Library*, CIO (June 3, 2025), https://www.cio.com/article/272361/infrastructure-it-infrastructure-library-itil-definition-and-solutions.html.

94. COBIT is an IT management framework created by the Information Systems Audit and Control Association to help organizations design, implement, and govern enterprise IT strategies effectively. Sarah K. White, *What Is Cobit? A Framework for Alignment and Governance,* CIO (June 12, 2023), https://www.cio.com/article/228151/what-is-cobit-a-framework-for-alignment-and-governance.html.

95. Elinda Kajo Meçe, Enida Sheme, Evis Trandafili, Carlos Juiz, Beatriz Gómez & Ricardo Colomo-Palacios, *Governing IT in HEIs: Systematic Mapping Review*, 11 BUS. SYS. RSCH. J. 93, 103 (2020).

96. *See, e.g.*, Daniël Smits & Jos van Hillegersberg, *The Continuing Mismatch Between IT Governance Theory and Practice: Results from a Systematic Literature Review and a Delphi Study with CIOs*, 24 J. MGMT SYS. 1 (2014).

97. Hicks et al., *supra* note 67, at 92.

98. Amrita Priyadarsini & Ajit Kumar, *A Literature Review on IT Governance Using Systematicity and Transparency Framework*, 24 DIGIT. POL'Y REGUL. & GOVERNANCE 309, 320 (2022).

99. *See* Matt Behrens, *How Iowa Centralized IT and Massively Overhauled State Systems*, GOV'T TECH. (May 7, 2025), ///how-iowa-centralized-it-and-massively-overhauled-state-systems; Liu et al., *supra* note 73, at 759.

100. Liu et al., *supra* note 73, at 780–81.

policies.[101] This lack of awareness can introduce barriers between strategic direction and implementation, which increases reliance on technology professionals for operational decisions that might otherwise be addressed at the governance level. This view is also held by campus stakeholders: faculty and librarians cite IT staff as having the most responsibility for assessing student technology needs and privacy concerns because they make decisions about technology procurement and support and also have the most technical knowledge about digital systems that involve student data.[102]

Despite serving important functions, chiefly ensuring maintenance of systems and cybersecurity, IT governance in higher education is characterized by a very limited engagement with students. Often, this practice is carried out by the broadly shared assumption that students do not care about their privacy. Outdated literature suggests that young people as "digital natives" that trade personal data for convenience and access.[103] However, more recent studies discredit this assumption, demonstrating that students are concerned about the widening scope and granularity of data collection, as well as what that data is used for by universities without their knowledge or consent.[104]

For students, technology governance remains a particularly opaque and exclusionary part of the university power structure. A 2024 EDUCAUSE survey found that 39 percent of higher education institutions reported no student involvement in technology governance processes, and an additional 35 percent characterized student participation as "ad hoc."[105] Just 6 percent of

101. *See* Ofir Turel, Peng Liu & Chris Bart, *Board-Level IT Governance*, 21 IT PRO. 58, 61 (2019).

102. Jones et al., *supra* note 63, at 336–37.

103. *See, e.g.*, Alessandro Acquisti & Ralph Gross, *Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook*, 4258 LECTURE NOTES COMPUT. SCI. 36, 37, 53–54 (2006); Susan B. Barnes, *A Privacy Paradox: Social Networking in the United States*, 11 FIRST MONDAY (Sep. 4, 2006), https://doi.org/10.5210/fm.v11i9.1394; Danah Boyd & Eszter Hargittai, *Facebook Privacy Settings: Who Cares?*, 15 FIRST MONDAY (July 27, 2010), https://doi.org/10.5210/fm.v15i8.3086.

104. *See, e.g.*, Lynne D. Roberts, Joel A. Howell, Kristen Seaman & David C. Gibson, *Student Attitudes Toward Learning Analytics in Higher Education: "The Fitbit Version of the Learning World"*, 7 FRONTIERS PSYCH., Dec. 19, 2016, at 5–8; JOSIE FISHER, FREDY-ROBERTO VALENZUELA & SUE WHALE, LEARNING ANALYTICS: A BOTTOM-UP APPROACH TO ENHANCING AND EVALUATING STUDENTS' ONLINE LEARNING (2014), https://ltr.edu.au/resources/SD12_2567_Fisher_Report_2014.pdf; Dirk Ifenthaler & Clara Schumacher, *Student Perceptions of Privacy Principles for Learning Analytics*, 64 EDUC. TECH. RSCH. & DEV. 923, 933–34 (2016); Park & Vance, *supra* note 23.

105. Ashley Caron, *QuickPoll Results: Positioning Higher Education IT Governance as a Strategic Function*, EDUCAUSE REV. (Feb. 21, 2024), https://er.educause.edu/articles/2024/2/educause-quickpoll-results-positioning-higher-education-it-governance-as-a-strategic-function.

universities reported that students serve as decision-makers.[106] This exclusion persists even as digital technologies become more central to students' academic and personal lives. The governance processes surrounding these systems are nontransparent to students, although they are most directly affected by data collection and digital transformations on campuses.[107]

Despite their limited power to shape these processes, students express clear desires for meaningful technology innovation, use, and governance. Students overwhelmingly believe that data collected about them should be de-identified and discarded once students exit the institution.[108] They expect universities to differentiate between the personal and academic spheres in data collection and use and strongly oppose the collection of biometric information.[109] Further, students perceive the sale of their data to any third parties as a violation of the trust they have in their universities.[110] Students clearly emphasize their ownership of their own data and their ability to regulate its collection, which are not recognized in turn by their universities.

Students also expect their data to be collected with an explicit plan for its use that improves the student experience. When institutions use student data for analytics, students expect such features to support their academic plans, organize their learning process, provide self-assessments, offer personalized recommendations, and create analyses of their learning.[111] These preferences are not inherently misaligned with university goals; on the contrary, they could help institutions design more effective, student-centered technologies. However, universities and vendors rarely engage students as co-creators and learning experts during the development or deployment of these systems, missing the opportunity to align data practices with the priorities of those most affected.

Given students' trust in universities to steward their data, and the growth of technology in universities as a powerful and largely unaccountable system of control, questions of student agency in technology and data governance become more important and urgent. How have we come to accept student exclusion from the governing processes that deeply impact their learning, privacy, and well-being? In the next Part, we explore more generally the historical and deeply entrenched exclusion of students from American

106. *Id.*

107. Madelyn Rose Sanfilippo, Noah Apthorpe, Karoline Brehm & Yan Shvartzshnaider, *Privacy Governance Not Included: Analysis of Third Parties in Learning Management Systems*, 124 INFO. & LEARNING SCIS. 326, 342 (2023).

108. Jones et al., *supra* note 25, at 499.

109. Park & Vance, *supra* note 23.

110. Jones et al., *supra* note 25, at 501, 503–04.

111. Clara Schumacher & Dirk Ifenthaler, *Features Students Really Expect from Learning Analytics*, 78 COMPUTS. HUM. BEHAV. 397, 404–05 (2018).

university governance, before offering an alternative governing and legal framework designed to address these exclusions.

IV. THE MARGINALIZATION OF STUDENTS FROM UNIVERSITY GOVERNANCE

Universities have become more technology-oriented and data-driven, at a time when students are increasingly marginalized or excluded from university governance. One scholar describes university governance as antidemocratic: "The power of rule is concentrated in a body whose members are neither selected by nor formally accountable to those over whom they rule."[112] Those with the power to rule describe their arrangement as one of shared governance, but such sharing usually only extends to administrators and faculty as students are given token representation at best in the boards and committees that govern these institutions.[113]

Even though students today do not hold meaningful governance roles, they historically were able to exercise power in universities. Early models of the university emphasized shared governance between students and faculty. Though this ideal eroded in the United States under the doctrine of *in loco parentis* ("in the place of parents"), it was partially revived during the wave of student protests in the 1960s and 1970s, when students won greater recognition and some formal roles in university decision-making. In recent decades, however, student influence has declined sharply, coinciding with the rise of managerialism and the rapid expansion of surveillance and data technologies on campus. The following Section traces this shifting relationship between students and institutional governance.

In medieval Europe, students shared governance power with faculty, municipal, and religious authorities at some early universities. The most notable example of student control emerged at the University of Bologna in the thirteenth century.[114] There, foreign students lacking the same legal protections as citizens organized themselves into protective guilds known as nations. As the city government's influence over the university increased,

112. TIMOTHY V. KAUFMAN-OSBORN, THE AUTOCRATIC ACADEMY: REENVISIONING RULE WITHIN AMERICA'S UNIVERSITIES 14–15 (2023).

113. *See, e.g.*, Josephine A. Boland, *Student Participation in Shared Governance: A Means of Advancing Democratic Values?*, 11 TERTIARY EDUC. & MGMT. 199, 200 (2005) ("While governance within higher education attracts increasing critical attention, participation of students has not featured prominently in these discussions.").

114. *See* SVEN STELLING-MICHAUD, L'UNIVERSITÉ DE BOLOGNE ET LA PÉNÉTRATION DES DROITS ROMAINS ET CANONIQUES EN SUISSE AU XIIIE ET XIVE SIÈCLES [THE UNIVERSITY OF BOLOGNA AND THE INFLUENCE OF ROMAN AND CANON LAW INTO SWITZERLAND IN THE 13TH AND 14TH CENTURIES] 26 (1955), cited in 1 A HISTORY OF THE UNIVERSITY IN EUROPE 213, 260 (H. De Ridder-Symoens & Walter Ruegg eds., 1988).

students consolidated these guilds into a student union with elected rectors, independent legal status, and its own statutes.[115] A council of student rectors and elected representatives from the nations served as the university's central governance structure.[116] This body exercised significant control over the faculty: electing teaching doctors annually, controlling their salary, and fining them for infractions, including deviation from student-approved curricula and poor lecture quality.[117] Students' control over faculty pay and their capacity for collective violence, such as arson, riots or building takeovers,[118] enabled them to enforce compliance with this structure despite opposition from faculty and city officials.[119] Although this extreme form of student power was later curtailed by the advent of salaried professorships funded by civic authorities, the model of governance at the University of Bologna marks a possible high point of student institutional control.[120]

The model of governance pioneered at Bologna served as a prototype for other Southern European universities. Within Italy, the University of Padua replicated the same model of complete student control.[121] Even though student power at this level was controversial in Italian society, there was broad acceptance of the principle that students should wield some degree of authority within the university.[122] Variants of student governance appeared at other Italian universities, including Perugia, Pisa, Florence, Pavia, Ferrara, Vicenza, Vercelli, and Piacenza, where students shared power with faculty and civic authorities.[123]

At French provincial universities around the same time, student power grew in tandem with opposition to ecclesiastical control over universities. French provincial universities fused elements of the Bolognese model with the hierarchical structure of the University of Paris, where students had virtually

115. *See* GUIDO ROSSI, "UNIVERSITAS SCHOLARIUM" E COMUNE (SEC. XII–XIV) ["UNIVERSITY SCHOLARSHIP" COMMUNE (12TH–14TH CENTURIES)] 175 (1955), cited in 1 A HISTORY OF THE UNIVERSITY IN EUROPE 85 (H. De Ridder-Symoens & Walter Ruegg eds., 1988).

116. STELLING-MICHAUD, *supra* note 114, at 123–24.

117. *See* Gaines Post, Masters' Salaries and Student-Fees in the Mediaeval Universities, 7 SPECULUM 181, 191–92 (1932).

118. Jonathan Davies, *Violence and Italian* Universities *During the Renaissance*, 27 RENAISSANCE STUD. 504, 508–09 (2013).

119. PEARL KIBRE, THE NATIONS IN THE MEDIAEVAL UNIVERSITIES 123–29 (1948).

120. Alan B. Cobban, *Medieval Student Power*, 53 PAST & PRESENT 28, 43–44, 65 (1971).

121. HASTINGS RASHDALL, THE UNIVERSITIES OF EUROPE IN THE MIDDLE AGES 9 (2010).

122. KIBRE, *supra* note 119, at 123–29.

123. RASHDALL, *supra* note 121, at 1–62.

no formal authority.[124] While the University of Paris remained controlled by faculty, student revolts across the provinces led to negotiated governance contracts that balanced the claims of faculty and students in university governance.[125] Student nations emerged at several universities, including Montpellier.[126] While students and faculty agreed that universities should be independent, faculty ultimately aligned themselves with ecclesiastical authorities rather than share institutional power. Despite this, student participation in governance persisted into the fifteenth century at universities in Aix, Poitiers, Valence, Nants, and Bourges while religious influence substantially receded.[127]

In Spain, some universities also emulated the University of Bologna's model. At the University of Salamanca, students formed their own council of elected rectors and formed nations.[128] At Valladolid, students occupied half of the seats on the university's governing council and held authority over the election of faculty to the other seats.[129] Other Spanish universities, including those at Lerida, Perpignan, and Huesca, incorporated similar elements of student governance to varying degrees.[130]

Ultimately, the model of the "masters' university" originating in Paris came to dominate the European university landscape. This model was premised on the idea that faculty held jurisdiction over academic matters, student discipline, and institutional governance.[131] It was adopted at English universities, including Oxford and Cambridge, as well as Northern European universities broadly.[132] Even in Italy and France, where some universities followed an alternative model, many institutions trended towards professionalized faculty governance and eroded student power by the end of the fifteenth century.[133]

The first American colleges modeled themselves after English universities. Colleges such as Harvard, Yale, and the College of William & Mary, founded in the years when large parts of America were a British colony, largely embraced the same governance structure, placing authority in the hands of

124. *See* V.R. Cardozier, *Student Power in Medieval Universities*, 46 PERS. & GUIDANCE J. 944, 944–45 (1968).

125. MARCEL FOURIER, LES STATUTS ET PRIVILÈGES DES UNIVERSITÉS FRANÇAISES DEPUIS LEUR FONDATION JUSQU'EN 1789 [THE STATUTES AND PRIVILEGES OF FRENCH UNIVERSITIES FROM THEIR FOUNDATION IN 1789] 189 (1890).

126. WALTER RUEGG, A HISTORY OF THE UNIVERSITY IN EUROPE 13 (Hilde De Ridder-Symoens ed., 1992).

127. RASHDALL, *supra* note 121, at 186.

128. KIBRE, *supra* note 119, at 156–57.

129. RASHDALL, *supra* note 121, at 69.

130. Cobban, *supra* note 120, at 56.

131. *Id.*

132. Cardozier, *supra* note 124, at 945, 947–48.

133. Cobban, *supra* note 120, at 65.

faculty and appointed boards.[134] However, the political break of the United States with England and its embrace of democratic ideals led to innovations in student governance.

The origins of formal student governance in the United States are somewhat contested. Some accounts trace the first student government to the College of William & Mary, which adopted a student-led honor code in 1736.[135] Others point to the University of Virginia, chartered in 1816 and founded by Thomas Jefferson. Jefferson, who had studied at William & Mary, designed the University of Virginia to include a student-led honor system and governance structure with an elective curriculum.[136]

The eighteenth and nineteenth centuries also saw the rise of a contrasting paternalistic approach to university administration, *in loco parentis*. The approach's origin is associated with William Blackstone's eighteenth-century *Commentaries on the Laws of England*.[137] The range of subjects covered by Blackstone's *Commentaries* is remarkable and its influence on the development of American law even more so. One chapter that continues to serve as a guiding hand in Americans' understanding of law and institutions focuses on the parent-child relationship. As a starting point, Blackstone explains: "The duty of parents to provide for the maintenance of their children, is a principle of natural law."[138] Even as some might disagree with its source in the natural law, Blackstone's account of the parental duty is rather uncontroversial. What followed were deeply contested claims, including that children are denied any other rights than those "given them by favour of their parents, or the positive constitutions of the municipal law" and that children "owe subjection and obedience [to their parents] during [their] minority, and honor and reverence ever after."[139]

---

134. *See* FREDERICK RUDOLPH, THE AMERICAN COLLEGE AND UNIVERSITY: A HISTORY 3, 13–16 (1990).

135. *Student Accountability and Restorative Practices: Honor Code & Honor Councils*, WM. & MARY, https://www.wm.edu/offices/communityvalues/sarp/honorcodeandcouncils/ (last visited Sep. 5, 2025); *see* HARRY C. MCKOWN, THE STUDENT COUNCIL 11 (1944); Walter P. May, *The History of Student Governance in Higher Education*, 28 COLL. STUDENT AFFS. J. 207, 210 (2010).

136. Others point to the University of Virginia, chartered in 1816 and founded by Thomas Jefferson. Jefferson, who had studied at William & Mary, designed the University of Virginia to include a student honor pledge and elective curriculum. MERRILL D. PETERSON, THOMAS JEFFERSON AND THE NEW NATION: A BIOGRAPHY 919 (1975); Coy Barefoot, *The Evolution of Honor: Enduring Principle, Changing Times*, VA. MAG. (Feb. 18, 2008), https://uvamagazine.org/articles/the_evolution_of_honor#1825.

137. *See generally* WILLIAM BLACKSTONE & THOMAS M. COOLEY, COMMENTARIES ON THE LAWS OF ENGLAND (Chicago, Callaghan & Cockcroft 1871).

138. *Id.* at 446.

139. *Id.* at 446, 453.

Centrally relevant to the student-university relationship is Blackstone's assertion that "[the father] may also delegate part of his parental authority, during his life, to the tutor or schoolmaster of his child; who is then *in loco parentis*."[140] Included in that delegation is the power "of restraint and correction, as may be necessary to answer the purposes for which he is employed."[141] Aside from the outdated notion of patriarchy that permeates the writing, Blackstone's *Commentaries* raise questions about the appropriateness and extent of this delegated authority to schools and agency denial to parental offsprings. Does the school's power include the power to make and apply whatever rules it deems proper? Should students have any role in the making of the rules? Do students have the obligation to obey arbitrary rules and applications? Can students be punished or expelled for alleged rules violations without process? What duty of maintenance and care does a school owe to the student?

Although Blackstone's focus was on the relationship between primary schools and children, these questions became central to litigation disputes between universities and students as universities sought to assert paternalistic authority over the students they enrolled. In the nineteenth and early twentieth centuries, American courts answered these questions in ways that reaffirmed colleges' and universities' parental dominion over students applying the doctrine of *in loco parentis*. In 1913, the Kentucky Supreme Court wholly embraced Blackstone's *in loco parentis* view of the student-university relationship in a decision that both exemplified the prevailing legal sentiment and subsequently influenced other courts and legal institutions. The case of *Gott v. Berea College* centered around a dispute involving the private college's application of a student manual rule that forbade students from "entering any 'place of ill repute, liquor saloons, gambling houses,' etc."[142] Students played no role in the rulemaking process. Nonetheless, for the court, that exclusion had no relevance for the student's obligation to obey the rule. What did have relevance was the parents' decision to affiliate their child with the college and the college's decision to admit the child. From those decisions arose the student obligation to "abide by and conform to the rules and regulations provided by the governing authorities of the college for the conduct of the students . . . upon pain of dismissal."[143] The court's reasoning about the student-college relationship followed from its understanding of the college's standing as "in loco parentis concerning the physical and moral welfare and

140. *Id.* at 453.
141. *Id.*
142. 161 S.W. 204, 205 (Ky. Ct. App. 1913).
143. *Id.* at 206.

mental training of the pupils."[144] That recognition of a quasi-parental status gave Berea College the authority to "make any rule or regulation for the government or betterment of their pupils that a parent could for the same purpose."[145] If the student thinks that the rules and their application are unwise or unjust, it is for the parents, not the courts, to intervene. In the absence of such parental intervention, the court explained that the school, "like a father may direct his children, [may] . . . direct their students what to eat and where they may get it, where they may go, and what forms of amusement are forbidden."[146]

In the decades that followed, the doctrine of *in loco parentis* shielded universities and colleges from judicial review of their rules and enforcement actions against students.[147] Exempt from external legal scrutiny, many of these rules became quite intrusive, restricting student autonomy and denying students the opportunity to exercise constitutionally recognized rights. As one scholar recounts, students were dismissed for smoking, "display[ing] behaviors 'unbecoming a typical Syracuse girl,' . . . skipping chapel, for consciously objecting to military drill, for writing private letters critical of the administration, and for marrying in a civil rather than a religious ceremony."[148] Students ordinarily had no say in the adoption of the rule to which they were subjected and were often denied due process in the colleges' enforcement of the rules. Instead, as another scholar explains:

> The validity of a college rule restricting the way in which students might spend their time or money, places they might go, people with whom they might associate, where they might live, etc., came to be tested by analogy; could a parent have maintained a similar rule in the supervision of his offspring at home?[149]

Student governance regained momentum as the Great Depression and international tensions sparked increased political engagement.[150] After World War II, returning veterans, empowered by the G.I. Bill, demanded greater

144. *Id.*

145. *Id.*

146. *Id.* at 207.

147. *See, e.g.*, Brian Jackson, *The Lingering Legacy of "*In Loco Parentis"*: An Historical Survey and Proposal for Reform*, 44 VAND. L. REV. 1135, 1147 (1991) ("The use of *in loco parentis* amounted to blanket judicial approval for all disciplinary actions against students.").

148. Christopher P. Loss, *Institutionalizing* In Loco Parentis *after* Gott v. Berea College *(1913)*, 116 TCHRS. COLL. REC. 1, 5 (2014).

149. William W. Van Alstyne, *The Tentative Emergence of Student Power in the United States*, 17 AM. J. COMPAR. L. 403, 406 (1969).

150. *See* Philip G. Altbach, *Student Politics: Activism and Culture*, *in* 18 INTERNATIONAL HANDBOOK OF HIGHER EDUCATION 329, 337 (James J.F. Forest & Philip G. Altbach eds., 2007).

student services and increased student participation in university governance.[151] By the end of the 1950s, student opposition to *in loco parentis* doctrine was growing as cultural norms shifted.[152] These developments paved the way for the sustained student movements of the 1960s.

Student opposition morphed into legal challenges of the status quo, including those rules and regulations adopted and enforced by colleges and universities. In 1961, approximately thirty African Americans at the Alabama State College for Negroes entered a restaurant, sat at the whites-only lunch counter, and requested service. The restaurant refused, closed the lunchroom, and called the police, who ordered the students to leave. The college's president then expelled the students for violating the college's rule of conduct without any notice or hearing, and the students later challenged their expulsion in federal court.[153] The trial court in *Dixon v. Alabama State Board of Education* upheld the expulsion pursuant to *in loco parentis*, but the appellate court reversed. The Fifth Circuit Court of Appeals held that "[w]henever a governmental body acts so as to injure an individual, the Constitution requires that the act be consonant with due process of law."[154]

After *Dixon*, courts imposed constitutional limits on college and university enforcement rules with the shifting judicial understanding of the status of students playing a critically important role. Courts prohibited colleges from enforcing rules in a manner that violated students' due process, free speech, and freedom of assembly rights, among others.[155] Concurrent with this recognition of student rights against colleges' rule enforcement actions came an evolution in the student-university relationship. Rather than parent-child, the students and universities came to be understood as contracting parties.[156] The evolving understanding of the relationship corresponded with a changing legal conception of college-aged students. The twenty-sixth amendment to the U.S. Constitution ratified in 1971 lowered the voting age from twenty-one to eighteen, signaling a shift in the status of most college-aged students from child to adult.[157] According to Blackstone, the original proponent of *in loco parentis*,

151. *See* PHILIP G. ALTBACH, STUDENT POLITICS IN AMERICA: A HISTORICAL ANALYSIS 122 (1997).

152. *See* Helen Lefkowitz Horowitz, *The 1960s and the Transformation of Campus Cultures*, 26 HIST. EDUC. Q. 1, 25–27 (1986).

153. *See* Dixon v. Ala. State Bd. of Educ., 294 F.2d 150, 152 (5th Cir. 1961) (describing the background facts of the case).

154. *Id.* at 155.

155. Jackson, *supra* note 147, at 1150–51.

156. *See* WILLIAM A. KAPLIN, BARBARA A. LEE, NEAL H. HUTCHENS & JACOB H. ROOKSBY, THE LAW OF HIGHER EDUCATION 363–64 (6th ed. 2020) (explaining that after *Dixon*, "courts increasingly viewed students as contracting parties having rights under express and implied contractual relationships with their institutions").

157. U.S. CONST. amend. XXVI.

that change was significant. As he explained, "[t]he legal power of a father [and his delegate] . . . over the persons of his children ceases" at the point at which the law establishes as the age of majority.[158]

During the following decade, students mobilized in unprecedented numbers to challenge their institutions and the broader political establishment. The Civil Rights Movement, the Free Speech Movement at Berkeley, and protests against the Vietnam War—coupled with universities' repressive responses—sparked student desire for an established voice in institutional decision-making. Students demanded participation in hiring and curriculum design, board and disciplinary body representation, control over student fees, and greater independence for student governments.[159] They envisioned seats on governing boards as a solution to their status as outsiders to decision-making at their institutions.[160]

The 1970s marked the culmination of student activism in the preceding decade as students gained institutional roles and voting rights.[161] Within universities, students successfully secured seats on their boards of trustees.[162] The number of student board seats would continue to increase thereafter for decades.[163] The lowering of the voting age to eighteen in the federal constitution allowed students to participate directly in state elections. With newfound electoral power, students helped to legally enshrine student governance expansions, such as student board positions, at public universities.[164] During this period, activists increasingly emphasized electoral organizing alongside campus governance. However, as the Vietnam War ended, student engagement in campus activism waned.[165] Many students moved their political involvement off-campus, leading to a gradual decline in

158. BLACKSTONE & COOLEY, *supra* note 137, at 453.

159. Angus Johnston, *Student Protests, Then and Now*: *From 'Hey, Hey, LBJ!' to 'Black Lives Matter!'*, CHRON. HIGHER EDUC. (Dec. 11, 2015), https://www.chronicle.com/article/student-protests-then-and-now/.

160. *See* CHRISTOPHER P. LOSS, BETWEEN CITIZENS AND THE STATE: THE POLITICS OF AMERICAN HIGHER EDUCATION IN THE 20TH CENTURY 165–214 (2014); Philip G. Altbach, *Perspectives on Student Political Activism*, 25 COMPARATIVE EDUC. 97, 100–02 (1989); Jon Lozano, *Bridging the Divide: Exploring the Connections Between Student Governments and Higher Education Governing Boards*, 45 STUD. HIGHER EDUC. 1878, 1879 (2020).

161. Altbach, *supra* note 150, at 337.

162. Lozano, *supra* note 160, at 1879; *see* Ray Allen Muston, Policy Boards and Student Participation 39 (1970) (Ph.D. Dissertation, Indiana University) (ProQuest).

163. *Student Trustees*, ASS'N OF GOVERNING BDS., http://agb.org/briefs/student-trustees [https://web.archive.org/web/20160416031054/https://www.agb.org/briefs/student-trustees]; Lozano, *supra* note 160, at 1879.

164. Lozano, *supra* note 160, at 1878–79; *see* Johnston, *supra* note 159.

165. Johnston, *supra* note 159.

student participation despite the achievement of structural gains in student governance.[166]

Following this development, students began to redefine their relationship with universities through a consumerist lens, shifting the student-university relationship to contracting parties: students now were not just participants in a scholarly endeavor, but also buyers of an educational product.[167] Rising costs, institutional competition, economic uncertainty, and student loan debt reinforced the idea that universities were offering a service and students were their paying customers.[168]

Though the terms of the student-university relationship had evolved, the nature of the relationship remained hierarchical, exclusionary, and undemocratic. Nominally described as an agreement between the student and the university, the "contract" operated more as one of adhesion than a mutual meeting of the mind.[169] Universities continued to establish the rules with minimal student input and the students consented to them as a condition to enrolling. Thus, when it came to the making of the rules in colleges and universities, the spirit of *in loco parentis* lived on.

The persistent attachment of *in loco parentis* to university governance was far from a foregone conclusion in the 1960s era of student activism. In fact, student challenges to exclusion from governance advanced in parallel to their efforts to secure protections for their constitutional rights from colleges and universities' rules enforcement. Faculty who had separately participated in a long-term struggle with university administrators and states to secure a role for themselves in university governance initially advocated for student participation in governance, especially in the 1960s.

For example, in 1967, the American Association of University Professors issued a joint statement advocating for a more inclusive university governance process. The purpose of the university, the statement began, was to transmit knowledge, pursue truth, develop students, and promote the general well-

166. *See* Altbach, *supra* note 150, at 337.

167. Anthony D. Plunkett, *A's for Everyone: The Effect of Student Consumerism in the Post-Secondary Classroom*, 19 QUALITATIVE REP. 1 (2014); Joan S. Stark, *The Many Faces of Consumerism*, 1976 NEW DIRECTIONS FOR HIGHER EDUC. 89, 91–92.

168. *See* CAITLIN ZALOOM, INDEBTED: HOW FAMILIES MAKE COLLEGE WORK AT ANY COST 16–19 (2019); Elizabeth Popp Berman & Abby Stivers, *Student Loans as a Pressure on U.S. Higher Education*, 46 RSCH. SOCIO. ORGS. 129, 144–46, 151 (2016); Michael Mulnix, *College Students as Consumers: A Brief History of Educational Marketing*, 2 J. MKTG HIGHER EDUC. 123, 137–38 (1990).

169. *See* Van Alstyne, *supra* note 149, at 411 (explaining that the terms of the student-college contracts "were nonnegotiable, many were vague phrased, and unilateral authority respecting their revisions, interpretation, and administration was reserved to the college").

being of society.[170] "[A]ll members of the academic community" share in this responsibility of establishing the conditions necessary to advance these purposes.[171] It is therefore a duty of "each college and university . . . to develop policies and procedures which provide and safeguard this freedom."[172] The making of such policies and procedures, the statement continues, should involve "the broadest possible participation of the members of the academic community."[173] That includes administrators, faculty, and students. As part of their governance role, students should not only be free "to express their views on issues of institutional policy and on matters of general interest to the student body."[174] They should also "have clearly defined means to participate in the formulation and application of institutional policy affecting academic and student affairs."[175]

In a subsequent statement that same year from the American Association of University Professors on Government of Colleges and Universities, the professors acknowledged that "students do not . . . have a significant voice in the government of colleges and universities."[176] But it called for institutes of higher education to involve those students who "desire to participate responsibly in the government of the institutions they attend . . . in the affairs of their college or university."[177]

The faculty's advocacy for students' participation in governance was, however, expressed more in words than in sustained actions as faculty remained divided on the subject. As a contemporaneous scholar explained, many faculty members "resent new demands for power sharing by students whose transient status, marginal educational expertise, and shorter perspectives appear to provide them with less than wholly attractive qualifications for the job."[178] Those understanding of student limitations are also noted in the current AAUP's statement on Government of Colleges and Universities that describes the obstacles to student participation in governance as "large and should not be minimized."[179] It calls for respecting opportunities for students:

---

170. Am. Ass'n of Univ. Professors, *Joint Statement on Rights and Freedoms of Students*, 54 AAUP BULL. 258 (1968).

171. *Id.*

172. *Id.*

173. *Id.* at 258–59.

174. *Id.* at 260.

175. *Id.*

176. Am. Ass'n of Univ. Professors, *Statement on Government of Colleges and Universities*, 52 AAUP BULL. 375, 375 (1966).

177. *Id.* at 379.

178. Van Alstyne, *supra* note 149, at 405.

179. Am. Ass'n of Univ. Professors, *supra* note 176, at 379.

> (1) to be listened to in the classroom without fear of institutional reprisal for the substance of their views, (2) freedom to discuss questions of institutional policy and operation, (3) the right to academic due process when charged with serious violations of institutional regulations, and (4) the same right to hear speakers of their own choice as is enjoyed by other components of the institution.[180]

Notably missing from the AAUP-endorsed governance allowances is the power of students to participate in the actual making of rules. Students have not been entirely excluded from these more substantive governing processes, but their role is very limited. Acknowledging the many variations, one scholar offers a general framework for how governance works at the university level. At the top of the pyramid is a governing board who has the "fiduciary duty . . . to advance the mission of the college or university over which it presides."[181] The board selects a president "to whom significant powers are delegated."[182] The president then hires officers and other administrators to carry out its delegated powers. Operating below this administrative level are the faculty of the various departments, who set policy regarding faculty hiring, the academic program, and curriculum. This structure of "shared governance" includes only a small number of entry points for student participation, such as participation in discussions on faculty hires.[183]

Today, many Boards of Regents or Governors for public colleges and universities do have a state legal mandate to involve students. In fact, our research shows that the laws in forty-two of the fifty states grant some participatory role to students on these bodies.[184] In those states, however, students have either minimal representation, no voting power, or both. The California State University and the National Association of System Heads conducted a survey of twenty-five state system governing boards and found that students had voting power on eleven of the boards, representation but no voting power on ten of the boards, and no representation at all on four of the boards.[185] On boards in which students had voting power, there was usually

180. *Id.*

181. KAUFMAN-OSBORN, *supra* note 112, at 12.

182. *Id.*

183. *Id.* at 12–13.

184. We examined publicly available websites describing the operation of state university boards in all 50 states to identify provisions for student participation on state university higher education boards. From this research, we found that 42 states provide for some student role in their state university higher education boards (on file with authors).

185. CAL. STATE UNIV. & NAT'L ASS'N OF SYS. HEADS, CHARACTERISTICS OF PUBLIC SYSTEM BOARDS IN US POSTSECONDARY EDUCATION (Nov. 2022).

only one student representative.[186] Students, therefore, have little or no input on the selection of the president or the administrators he appoints. Faculty in the departments that set faculty hiring, academic, and curricular policies usually do so through a committee structure in which students are either excluded or only given nominal representation with little to no power to influence decisions. Structurally, students serve as focus groups or consultants, rather than decision-makers. In a real sense, shared governance is really governance shared between administrators and faculty only.

When considering governance surrounding technology, there is a certain irony associated with student exclusion from university decision-making bodies. Studies consistently show that young people are the earlier adopters of technology and the more frequent users of the tools.[187] This is a trend that has continued with the emergence of AI.[188] For example, students use chatbots to help them understand complex concepts, fix grammar and language in essays, or study for quizzes. Further, students are much more likely than faculty and administrators to use and be subject to technology as part of participating in university life, including as part of their learning experience.[189] Good governance and rulemaking are typically associated with evidence-based information about the thing or activity being regulated. That makes students particularly well-positioned to actively participate in technology governance and help shape decisions about technology policy in college and university settings. In the next Part, our focus is on the problems associated with student exclusion from technology governance. We then draw from theory and practice to make the case for greater democratization in university technology governance.

186. *Id.* at 6 tbl.A1 (showing that students had more than one representative on only four of the eleven boards in which they had voting power).

187. *See, e.g.*, Michelle Faverio, *Share of Those 65 and Older Who are Tech Users Has Grown in the Past Decade*, PEW RSCH. CTR. (Jan. 13, 2022), https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/ ("Younger adults are often more likely than their elders to bee earlier adopters of innovations.").

188. *See, e.g.*, Courtney Gregoire, *Increased Uptake of Generative AI Technology Brings Excitement and Highlights the Importance of Family Conversations About Online Safety, Says New Research from Microsoft*, MICROSOFT ON THE ISSUES (Feb. 5, 2024), https://blogs.microsoft.com/on-the-issues/2024/02/05/generative-ai-online-safety-day-global-survey/ (finding young adults to be "the most active users and experimenters" with AI).

189. *See, e.g.*, Lasha Labadze, Maya Grigolia & Lela Machaidze, *Role of AI Chatbots in Education: Systematic Literature Review*, 20 INT'L J. EDUC. TECH. HIGHER EDUC., Oct. 31, 2023, at 10.

V. REMEDYING THE PROBLEM OF STUDENT EXCLUSION FROM UNIVERSITY TECHNOLOGY GOVERNANCE

Students' desires for technology use are increasingly at odds with the prevailing institutional goals. Students overwhelmingly express a preference for systems that support their life on campus and learning experience, protect their privacy, and discard personal data upon graduation.[190] In contrast, universities and the third-party vendors they hire often retain student data indefinitely, repurposing it to make individualized predictions and interventions without meaningful student input. These practices prioritize institutional efficiency, liability management, and revenue generation over student autonomy. For example, rather than de-identify data, institutions apply AI and predictive analytics to nudge students toward efficient educational paths, flag at-risk behaviors, or determine access to resources, effectively narrowing student choice. Despite widespread student discomfort with biometric tracking, several universities collect biometric data, such as facial images, keystroke dynamics, and emotion recognition outputs.[191] These developments illustrate a deepening surveillance infrastructure that contrasts sharply with student preferences for clear boundaries between academic and personal spheres.

As the deployment of digital technologies that are based on pervasive student data collection coincides with a historic low of student-led governance at universities and a siloing of technology governance on campuses, three long-term implications on students crystallize.

First, the pervasive extraction of data from students is socializing them into passive roles within algorithmic systems while generating durable digital footprints with lasting implications for their future. From search activities and academic performance to personal habits and biometrics, universities collect continuous, granular data on students.[192] This normalizes an environment of surveillance, habituating students to constant monitoring and management.[193] Students are forced to relinquish their privacy as a condition of accessing a resource: in this case, education. We further know that surveillance is directly harmful to students, especially marginalized ones, as they already are more

190. Kyle M. L. Jones, Abigail Goben, Michael R. Perry, Mariana Regalado, Dorothea Salo, Andrew D. Asher, Maura A. Smale & Kristin A. Briney, *Transparency and Consent: Student Perspectives on Educational Data Analytics Scenarios*, 23 LIBRS. & ACAD. 485, 501–02 (2023).

191. *See* Danielle Keats Citron, *The Surveilled Student*, 76 STAN. L. REV. 1439, 1455–62 (2024).

192. *Id.*

193. *Id.* at 1459.

likely to experience surveillance in their daily lives.[194] University data collection and utilization practices condition students to accept surveillance and data extraction in the future. The university thus becomes a site where students are not only surveilled but actively trained to accept datafication and prediction as a prerequisite to living the life of a student, and of entering into the contract-based student-university relationship.

Students today will graduate into a labor market where productivity is increasingly quantified to improve efficiency and enhance profits.[195] University experiences in this context double as training in algorithmic subjection as students internalize the logic of being constantly extracted from and assessed by opaque digital systems and of being excluded from deliberations about the socio-technical outfit of a cornerstone institution like a workplace.[196]

Moreover, the comprehensive data generated during a student's time at university does not just disappear upon graduation. This data is retained and used by the university and third-party vendors. These digital profiles may later be repurposed in ways that students never anticipated, affecting employment prospects, eligibility for immigration visas, access to credit, or insurance premiums. Universities and third party technology vendors already collaborate closely with industry and potential employers and are seeking to expand these partnerships.[197] Some digital technologies already in use at universities are explicitly oriented toward career planning and readiness, collecting data that

194. *See* Pokorny et al., *supra* note 47, at 14–17; Barton Gellman & Sam Adler-Bell, *The Disparate Impact of Surveillance*, CENTURY FOUND. (Dec. 21, 2017), https://tcf.org/content/report/disparate-impact-surveillance/.

195. IFEOMA AJUNWA, THE QUANTIFIED WORKER: LAW AND TECHNOLOGY IN THE MODERN WORKPLACE 36–37 (2023).

196. *See, e.g.*, Wharton Staff, *Your Data Is Shared and Sold ... What's Being Done About It?*, KNOWLEDGE AT WHARTON (Oct. 28, 2019), https://knowledge.wharton.upenn.edu/article/data-shared-sold-whats-done/; *see also* Ulises A. Mejias & Nick Couldry, *Datafication*, 8 INTERNET POL'Y REV. 1, 3–4 (2019), https://policyreview.info/concepts/datafication; Jodi Kantor & Arya Sundaram, *The Rise of the Worker Productivity Score*, N.Y. TIMES (Aug. 14, 2022), https://www.nytimes.com/interactive/2022/08/14/business/worker-productivity-tracking.html; Ifeoma Ajunwa, *Algorithms at Work: Productivity Monitoring Applications and Wearable Technology as the New Data-Centric Research Agenda for Employment and Labor Law*, 63 ST. LOUIS U. L.J. 21, 33–34 (2018); Madeline Yingling, *Consumers Sue Amazon Over Alleged Tracking of Sensitive Data*, JURIST NEWS (Jan. 30, 2025), https://www.jurist.org/news/2025/01/consumers-sue-amazon-over-alleged-tracking-of-sensitive-data/.

197. Jan Lynn-Matern, *Mass Collaboration Between Employers and Universities is the Future of Higher Education | Part 1—Why Are We Investing in this Space?*, MEDIUM (Apr. 30, 2020), https://medium.com/emerge-edtech-insights/mass-collaboration-between-employers-and-universities-is-the-future-of-higher-education-part-1-ed840467bfd5; Natalia Kucirkova, *A Partnership Industry for Impactful Ed-Tech*, STAN. SOC. INNOVATION REV. (Apr. 22, 2024), https://ssir.org/articles/entry/ed-tech-partnership-industry; *see* Justin Ménard, *Unveiling the Power of Business Partnerships in EdTech*, LISTEDTECH (Feb. 28, 2024), https://listedtech.com/blog/unveiling-the-power-of-business-partnerships-in-edtech/.

may alter employers' interest.[198] Analytics companies have even pitched universities on the idea of tracking students after graduation as a part of a broader human datafication vision.[199] As universities have already stated the intention to diversify their revenue streams through student data, these possibilities are even more concerning.[200] Without safeguards or informed consent, universities and vendors could easily justify sharing student data under the guise of convenience and student support.

These developments raise significant concerns about fairness and bias in algorithmic modeling. As has already occurred in the hiring context, the creation of large-scale datasets and their use in algorithmic decision-making at universities can reproduce existing inequities.[201] At Georgia State University, for instance, a predictive analytics program disproportionately steered students of color toward lower-paying majors.[202] If this data and modeling continue to follow students even beyond the university, colleges risk harming students further. Many students are not even aware that their educational data can live on in such forms, and that their "data double" created during college could follow them for the rest of their lives.[203] These risks highlight the asymmetry of power in campus data practices: universities and vendors reap immediate insights and cost savings, while students bear potentially lifelong consequences of a far-reaching digital footprint.

Second, the exclusion of students from governance directly undermines their agency and self-determination as digital technologies violate their intellectual privacy.

Students in higher education possess a distinct form of agency, generally understood as the right to participate in and co-decide on matters that affect

198. *Workforce EdTech Tools*, EDTECH CENTER, https://workforceedtech.org/tools/ (last visited Aug. 17, 2025); Felicity Cartwright, *Navigating Career Paths with Innovative Ed-Tech Tools*, MENTORING TRENDS MEDIA BLOG (May 5, 2025), https://www.mentoring-trends.com/blog/navigating-career-paths-with-innovative-ed-tech-tools.

199. Jeffrey R. Young, *How Tech Companies Are Selling Colleges on Mass Data Collection*, EDSURGE (Oct. 18, 2019), https://www.edsurge.com/news/2019-10-18-how-tech-companies-are-selling-colleges-on-mass-data-collection.

200. Reisman, *supra* note 1, at 566–68; *see* WEINBERG, *supra* note 4, at 57–58.

201. Jeffrey Dastin, *Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women*, REUTERS (Oct. 10, 2018), https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/.

202. *See* MPR News Staff, *Under a Watchful Eye: How Colleges are Tracking Students to Boost Graduation*, MPR NEWS (Apr. 14, 2020), https://www.mprnews.org/story/2020/04/14/apm-reports-under-a-watchful-eye.

203. Mark Andrejevic & Kelly Gates, *Big Data Surveillance: Introduction*, 12 SURVEILLANCE & SOC'Y 185, 191 (2014).

themselves and others within university governance structures.[204] This agency legitimizes their inclusion in decision-making processes, even when their formal power remains limited. Students are active contributors to the academic, social, and political life of the university.[205] A broader argument for the agency of students can be made beyond the institutional context. Students frequently organize as a distinct political group, including in student labor unions and protests such as the 2012 Quebec student protests, the Million Student March, and the #FeesMustFall movement.[206] They have also played prominent roles in larger social movements, including gun violence prevention and Black Lives Matter.[207] These patterns highlight students' capacity for collective action and suggest that they constitute a unique political constituency with enduring governance aspirations. Further, if student status is a transitory yet common stage of citizenship, conceptualizing student governance as a distinct level of government offers a more comprehensive understanding of their role in public life.[208]

Yet despite their demonstrated capacity for meaningful political engagement, students are routinely denied agency within the governance structures of higher education. Most institutional decision-making remains concentrated in the hands of administrators and governing boards, with

204. Manja Klemenčič, *The Key Concepts in the Study of Student Politics and Representation in Higher Education*, *in* THE BLOOMSBURY HANDBOOK OF STUDENT POLS. AND REPRESENTATION IN HIGHER EDUC. 7, 10–19 (Manja Klemenčič ed., 2024).

205. *See* Justin Patrick, *Student Leadership and Student Government*, 7 RSCH. EDUC. ADMIN. & LEADERSHIP 1, 20–21 (2022).

206. *See* ROUTLEDGE HANDBOOK OF THE SOCIOLOGY OF HIGHER EDUCATION 95–97 (James E. Côté & Sarah Pickard eds., 2d ed. 2022); Parbudyal Singh, Deborah M. Zinni & Anne F. MacLennan, *Graduate Student Unions in the United States*, 27 J. LAB. RSCH. 55, 56–60 (2006); Leila Lemghalef, *Big Montreal March Marks 100 Days of Student Anger*, REUTERS CAN. (May 22, 2012), https://web.archive.org/web/20141021021206/http://ca.reuters.com/article/domesticNews/idCABRE84H12620120522; MALOSE LANGA, SANDILE NDELU, YINGI EDWIN, MUSAWENKOSI MALABELA, MARCIA VILAKAZI, OLIVER METH, GODFREY MARINGIRA, SIMBARASHE GUKURUME & MUNEINAZVO KUJEKE, CENTRE FOR THE STUDY OF VIOLENCE AND RECONCILIATION, #HASHTAG: AN ANALYSIS OF THE #FEESMUSTFALL MOVEMENT AT SOUTH AFRICAN UNIVERSITIES 6 (2017).

207. *See* ADAM FLETCHER, MEANINGFUL STUDENT INVOLVEMENT: GUIDE TO STUDENTS AS PARTNERS IN SCHOOL CHANGE 15 (2d ed. 2005); MARK EDELMAN BOREN, STUDENT RESISTANCE: A HISTORY OF THE UNRULY SUBJECT 1–2 (2001); Klemenčič, *supra* note 204, at 407; Altbach, *supra* note 150, at 335; Michael A. Goodman, *Openly Gay Undergraduate Men in Student Government: Out, Visible, and Elected*, 15 J. DIVERSITY HIGHER EDUC. 766, 767–68 (2022); Emily Bent, *Unfiltered and Unapologetic: March for Our Lives and the Political Boundaries of Age*, 11 JEUNESSE: YOUNG PEOPLE, TEXT, CULTURES 55, 60, 62 (2019); Christopher Rim, *How Student Activism Shaped The Black Lives Matter Movement*, FORBES (June 4, 2020), https://www.forbes.com/sites/christopherrim/2020/06/04/how-student-activism-shaped-the-black-lives-matter-movement/.

208. Patrick, *supra* note 205, at 21.

limited student representation and minimal influence over core issues such as budgeting, curriculum design, surveillance practices, and labor conditions. Even when students are granted formal roles such as board seats, these positions are often advisory. Students may be permitted to raise concerns but are rarely empowered to act on them.[209] This exclusion is reinforced by the perception of students as transient stakeholders, whose presence on campus is temporary and whose governance claims are seen as secondary to those of permanent staff.

The student representation allowed by universities often does not translate into an exercise of student agency, but is actually just a form of consultation or high-level participation.[210] For example, non-voting student members of boards of trustees are highly visible, but are denied any decision-making power, inherently limiting the power of the office of student trustee.[211] The exclusion of students from *technology* governance directly undermines their agency and self-determination in the digital sphere of campus life. Although students are the primary users of most campus technologies—generating the data that fuels analytics (from learning platforms to meal services) and bearing the consequences of its use—they are systematically denied a voice in decisions about the design, selection, or operation of these systems. This exclusion is perpetuated by the paternalistic posture of higher education institutions, which implies that institutions know better than students what technologies are appropriate, elevating the administrative interests of the university over student interests and student governance.[212] Without genuine efforts to address the imbalances of information, power, and agency between students and institutions, this approach to technology risks recreating *in loco parentis* in the modern day, with potentially far-reaching consequences.[213]

Students' intellectual privacy is uniquely threatened by their exclusion from decisions about the digital technologies that increasingly shape campus life. Intellectual privacy refers to the freedom to think, read, and speak in confidence, free from surveillance or interference.[214] Intellectual privacy is

209. Rebecca Freeman, *Is Student Voice Necessarily Empowering? Problematising Student Voice as a Form of Higher Education Governance*, 35 HIGHER EDUC. RSCH. & DEV. 859, 860–61 (2016).

210. SARAH K. ELFRETH, THE YOUNG GUARDIANS: STUDENTS AS STEWARDS OF THE PAST, PRESENT, AND FUTURE OF AMERICAN HIGHER EDUCATION 2–3 (Cooper Anderson & Matt Strauch eds., 2011); Lozano, *supra* note 160, at 1878, 1884–85.

211. Lozano, *supra* note 160, at 1884–85.

212. *See* Jeffrey Alan Johnson, *Ethics and Justice in Learning Analytics*, 2017 NEW DIRECTIONS FOR HIGHER EDUC. 77, 79 (2017).

213. *See* Paul Prinsloo & Sharon Slade, *Student Vulnerability, Agency, and Learning Analytics: An Exploration*, 3 ETHICS & PRIV. LEARNING ANALYTICS 159, 165–66, 178 (2016).

214. NEIL RICHARDS, INTELLECTUAL PRIVACY: RETHINKING CIVIL LIBERTIES IN THE DIGITAL AGE 11 (2015).

foundational to the development of independent thought, creativity, and dissent in a democratic society.[215] Under constant surveillance and the fear that today's data footprints will shape tomorrow's AI systems, students may self-censor their explorations, hesitating to read controversial materials, discuss sensitive topics online, or take intellectual risks, for fear that their digital traces could be misinterpreted or later used against them.

Students' lack of intellectual privacy runs directly counter to the purportedly democratic missions of universities. American universities identify as bastions of democracy, justifying their existence and decisions to stakeholders and the public by appealing to their role in producing informed citizens, enhancing political participation, and advancing social mobility.[216] However, for students still forming world views, intellectual privacy is especially necessary for civic and intellectual development.[217] It enables civic and intellectual development by creating space to explore, critique, and construct new ideas without fear of institutional surveillance. Accordingly, intellectual privacy for students is necessary for universities to create environments in which deliberative democracy can be practiced.[218] Despite this, the current state of technology governance at universities shows that the democratic and educational mission of higher education is overshadowed by technology industry imperatives. The institutional incentives in a competitive, marketized higher education sector encourage administrations to adopt technology tools in pursuit of efficiency and control, even when these conflict with student autonomy and privacy.

---

215. Julie E. Cohen, *What Privacy Is For*, 126 HARV. L. REV. 1904, 1912–18 (2013); Neil M. Richards, *The Dangers of Surveillance*, 126 HARV. L. REV. 1934, 1945–53 (2013).

216. THOMAS L. PANGLE, THE ENNOBLING OF DEMOCRACY: THE CHALLENGE OF THE POSTMODERN AGE 181 (2021); RONALD J. DANIELS, WHAT UNIVERSITIES OWE DEMOCRACY 17–27 (2021); James Dean, *How Universities Can Help Strengthen Democracy*, CORNELL CHRON. (Sep. 16, 2024), https://news.cornell.edu/stories/2024/09/how-universities-can-help-strengthen-democracy; George F. Zook, *The President's Commission on Higher Education*, 33 BULL. AM. ASSOC. UNIV. PRESIDENTS 10, 15–17 (1947); Patricia McGuire, *Higher Education and the Defense of Democracy: Confronting the Ideology of Ignorance*, ACADEME MAG. (2025), https://www.aaup.org/article/higher-education-and-defense-democracy; Sjur Bergan, Ira Harkavy, Rita Hodges, Ronaldo Munck, Yadira Pinilla & Hilligje van't Land, *Higher Education Institutions Are Anchors for Democracy*, UNIV. WORLD NEWS (July 5, 2022), https://www.universityworldnews.com/post.php?story=20220705222834768.

217. Citron, *supra* note 191, at 1444; DANIELS, *supra* note 216, at 86–117; *see* Jonathan Koppell, *The Role of Universities in Shaping Democratic Values*, MONTCLAIR STATE UNIV. (July 8, 2024), https://www.montclair.edu/president/2024/07/08/the-role-of-universities-in-shaping-democratic-values/; Tony Gallagher, *The Democratic Imperative for Higher Education: Empowering Students to Become Active Citizens*, LIBERAL EDUC. (2021), https://www.aacu.org/liberaleducation/articles/the-democratic-imperative-for-higher-education.

218. Neil M. Richards, *Intellectual Privacy*, 87 TEX. L. REV. 387, 407–25 (2008).

Several justifications have been proffered for the exclusion of students from university governance. Some of those justifications can be principally derived from the *in loco parentis* account of students. For the Kentucky Supreme Court in *Gott v. Berea College*, the early twentieth-century case that famously embraced the *in loco parentis* doctrine, the students were "inexperienced country, mountain boys and girls of little means."[219] While colleges recognize that students come from a variety of socioeconomic backgrounds and cater to them accordingly, the *in loco parentis* account of students as being of little experience, maturity, and understanding continues to inform attitudes towards student involvement in university governance. As one example, the American Association of University Professors identifies as the obstacle to student participation in university governance their "inexperience, untested capacity, [and] transitory status which means that present action does not carry with it subsequent responsibility."[220]

These justifications for student marginalization from university governance are overbroad as they fail to account for students' unique capacities and experiences that could be particularly beneficial for governance decisions. When it comes to technology specifically, students' tendency to be earlier adopters and more active users can put them at a comparative advantage vis-à-vis faculty and administrators in constructing rules targeting technology.[221] But even if we assume that faculty's and administrators' capacity and experience puts them in a superior position to govern over students, decision-making theory suggests that excluding students from governance will lead to worse regulatory outcomes.

Theorists have long advocated for deliberative decision-making processes that are broadly inclusive.[222] In such processes, all members of a polity, stakeholders of an institution, or their representatives should participate in a process in which views or ideas are exchanged prior to a decision.[223] The ultimate decisions might be the product of a consensus that emerges among participants after deliberation or a decision that a majority reaches after a hearing and consideration of all the participants' views. There are two key elements to the deliberative process. First, it must be inclusive. And second, decisions should be the product of the considered views of the participants.

---

219. 161 S.W. 204, 206 (Ky. Ct. App. 1913).

220. Am. Ass'n of Univ. Professors, *supra* note 176, at 379.

221. *See* discussion *supra* notes 185–187.

222. *See, e.g.*, Joshua Cohen*, Deliberation and Democratic Legitimacy*, *in* DELIBERATIVE DEMOCRACY: ESSAYS ON REASON AND POLITICS (James Bohman & William Rehg, eds. 1997); AMY GUTMANN & DENNIS THOMPSON, WHY DELIBERATIVE DEMOCRACY? (2004); JOHN S. DRYZEK, FOUNDATIONS AND FRONTIERS OF DELIBERATIVE GOVERNANCE (2010).

223. *See, e.g.*, JAMES BOHMAN, PUBLIC DELIBERATION: PLURALISM, COMPLEXITY, AND DEMOCRACY 5–6, 35–36 (1996) (describing a deliberative process).

There are two principal arguments advanced for why deliberation is better than other decision-making processes. The first is that the deliberative decision-making process is more democratic. Describing consent as the core feature of democracy, James Bohman argues that "democracy implies public deliberation in some form."[224] "The deliberation of citizens," he continues, "is necessary if decisions are not to be merely imposed upon them."[225] For that deliberative process to be truly democratically legitimate, Bernard Manin, Elly Stein, and Jane Mansbridge argue, it must also be inclusive. Since "political decisions are characteristically imposed on *all*, it seems reasonable to seek, as an essential condition for legitimacy, the deliberation of *all* or, more precisely, the right of all to participate in deliberation."[226]

This argument for deliberation accords with the democracy-enhancing functions of the university. The primary purpose of American universities has been historically understood as the education of students and the creation of knowledge to produce an informed citizenry and a democratic, enlightened society.[227] Indeed, scholars identify institutions of higher education as essential for maintaining and strengthening democratic practices.[228] American universities, particularly public, liberal arts, and highly ranked institutions, explicitly frame their mission statements around such aims: advancing civic responsibility, fostering critical inquiry, and promoting social mobility.[229] Within this framework, the structure and organization of technology governance is not solely an operational concern, but a mechanism for aligning digital technologies and practices to the values that an institution wants to perpetuate.

Some might reject the relevance of this democratic justification for a more inclusive deliberative process. Some contest the claim that universities are, or should be, democratic institutions. As a descriptive matter, universities are not particularly democratic institutions. As described earlier, American universities in their current incarnation are rather hierarchical institutions in which students, for the most part, do not consent to their governors or the rules that

224. *Id.* at 4.

225. *Id.*

226. Bernard Manin, Elly Stein & Jane Mansbridge, *On Legitimacy and Political Deliberation*, 15 POL. THEORY 338, 352 (1987).

227. *See* ROBERT B. WESTBROOK, JOHN DEWEY AND AMERICAN DEMOCRACY 171–72 (1991); HERBERT CROLY, THE PROMISE OF AMERICAN LIFE 405 (Arthur M. Schlesinger, Jr. ed., 1965); Albert Castel, *The Founding Fathers and the Vision of a National University*, 4 HIST. EDUC. Q. 280, 281–82 (1964).

228. *See supra* note 216.

229. Citron, *supra* note 191, at 1455–56; Courtney H. Thornton & Audrey J. Jaeger, *Institutional Culture and Civic Responsibility: An Ethnographic Study*, 47 J. COLL. STUDENT DEV. 52, 63–64 (2006); Ira Harkavy, *The Role of Universities in Advancing Citizenship and Social Justice in the 21st Century*, 1 EDUC., CITIZENSHIP & SOC. JUST. 5, 7–12 (2006).

they make.[230] And for reasons we also described earlier, some argue that universities are appropriately undemocratic.[231] If we assume universities are not democratic and that they should not be, a deliberative decision-making process for universities grounded in democracy becomes highly questionable.

There is, however, a second argument for deliberation relevant to universities that does not depend on them being democratic institutions. That argument for a deliberative process is that it leads to better decisions. As Manin et al. explain, in a deliberative process, there is an opportunity for individuals to "listen[] to arguments formulated by others," "broaden [their] point of view," and "become[] aware of things [they] had not perceived at the outset."[232] As Bernard Grofman and Scott Feld theorize, deliberation "provides information about who holds what preferences and diffuses information about why people hold the preferences that they do."[233] University governance often satisfies these requirements for deliberation. Many decisions are made through committee processes in which participants share information and provide arguments prior to a decision being made. Those participants tend to have experience in university governance, and some might even have expertise on the matters being addressed. The implicit assumption in university governance is that deliberative bodies comprised of individuals with deep experience and expertise will lead to optimal decisions.

Deliberative decision-making processes in universities through committees should lead to better decisions than those that arise from a more hierarchical decision-making structure. But contrary to what many might assume, a decision-making body comprised of a more homogeneous group of experienced and expert university administrators and faculty members will make worse decisions than a more diverse and inclusive deliberative decision-making process. To state differently, processes that include less experienced and less expert students will result in better decisions than those that exclude them.

In a seminal work, Lu Hong and Scott Page found that "a functionally diverse group whose members have less ability outperform a group of people with high ability."[234] These findings served as the basis for the Diversity Trumps Ability Theorem. James Surowiecki offered further support for the

230. *See* discussion *supra* notes 207–210.

231. *See* discussion *supra* notes 178–180.

232. Manin et al., *supra* note 226, at 352.

233. Bernard Grofman, David M. Estlund, Scott L. Feld & Jeremy Waldron, *Democratic Theory and the Public Interest: Condorcet and Rousseau Revisited*, 83 AM. POL. SCI. REV. 1317, 1333 (1989).

234. Lu Hong & Scott E. Page, *Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers*, 101 PNAS 16385, 16385 (2004).

theorem. He explains, "[d]iversity helps because it actually adds perspectives that would otherwise be absent and because it takes away, or at least weakens, some of the destructive characteristics of group decision making."[235] For the best decisions, "intelligence alone is not enough, because intelligence alone cannot guarantee you different perspectives on a problem."[236] Homogenous groups of experienced and expert people perform worse because the group members are too much alike as "each member is bringing less and less new information to the table."[237] When a group brings "new members into the organization, even if they're less experienced and less capable," the group becomes "smarter simply because what little the new members do know is not redundant with what everyone else knows."[238]

Hélène Landemore, in her critical book on *Democratic Reason*, labels as cognitive diversity the diversity that is critical for making better decisions in a deliberative process. Cognitive diversity denotes "a diversity of perspectives (the way of representing situations and problems), diversity of interpretations (the way of categorizing or partitioning perspectives), diversity of heuristics (the way of generating solutions to problems), and diversity of predictive models (the way of inferring cause and effect)."[239]

Such cognitive diversity is a feature of universities with students generally bringing different backgrounds, experiences, and capacities from those of administrators and faculty. Those differences that often serve as the main justification for *excluding* students from university governance should, according to deliberative decision-making theory, be a primary reason for *including* students in university governance. Students share perspectives, interpretations, heuristics, and predictive models that will often vary from those of administrators and faculty. A university governance process inclusive of students, administrators, and faculty is therefore likely to lead to better rules applicable to university stakeholders.

Although university governance processes currently tend to be quite exclusionary, the good news is that there are very few legal constraints on creating more inclusive university governance processes. And there are no legal constraints that we have identified that would limit opportunities for creating more inclusive university technology governance processes. For public

235. JAMES SUROWIECKI, THE WISDOM OF CROWDS: WHY THE MANY ARE SMARTER THAN THE FEW AND HOW COLLECTIVE WISDOM SHAPES BUSINESS, ECONOMIES, SOCIETIES, AND NATIONS 38 (2004).

236. *Id.*

237. *Id.*

238. *Id.*

239. HÉLÈNE LANDEMORE, DEMOCRATIC REASON: POLITICS, COLLECTIVE INTELLIGENCE, AND THE RULE OF THE MANY 102 (2017).

colleges and universities, state law governs the composition of university boards and the voting power of its members. To obtain the benefits of better policy from a diverse and inclusive decision-making process, it is critical that the board includes students. As noted above, state laws in forty-two of the fifty states provide for student participation.[240] Changes to laws in eight states to add student members could, under the deliberative decision-making model, improve those state boards' rules and regulations. But more is likely to be necessary.

Deep hierarchies in higher education that are likely related to outdated paternalistic views many faculty and administrators hold toward students could limit the deliberative benefits from a more cognitively diverse decision-making process. Further steps might therefore be necessary to ensure that the perspectives students bring from their different backgrounds and experiences are seriously considered and contribute to better decisions being made. First, state law should provide students with voting rights on the college and university boards. Currently, only twenty-seven of the forty-two states that provide for student representation on university boards give those students voting privileges. The choice to deny students voting power is a clear signal of their second-class status on the board, which reinforces paternalistic attitudes that lead to their views being marginalized or ignored in the decision-making process.

Second, state law should provide for a critical mass of student representation on the board. In the California State University study of university boards, most of the university boards with student representatives had only one student member.[241] Even with the vote, a single student member on boards ranging from seven to twenty-five members will find it challenging to overcome obstacles to their views being seriously considered. As one member, student voices can be more easily drowned out, and with one vote, the student choice will rarely be pivotal for a decision. Including a critical mass of students on boards provides for secondary and tertiary reinforcement of student views expressed during board deliberations, which adds to the likelihood that those views will be heard by other members. Including a critical mass would also make serious consideration of student views more likely because of the greater odds that students will be a pivotal vote on policy decisions.

In our research on U.S. state laws pertaining to university governance, we have not identified any legal constraints placed on university and department-level administrators regarding committee composition and member voting

240. *See supra* note 184.

241. *See* CAL. STATE UNIV. & NAT'L ASS'N OF SYS. HEADS, *supra* note 185, at 7–8.

power. But there may be obvious privacy or prudential concerns associated with placing students on certain committees, such as student disciplinary committees. However, a distinction should be drawn between student exclusions from committees that are based on prudential and privacy reasons and those exclusions from committees that are based on longstanding paternalistic views regarding the competence and capacity of students. As deliberative decision-making theory shows, it is the different backgrounds and experiences of students that add to the quality of decisions. Paternalistic-based student exclusions from committees should therefore be replaced by an orientation toward inclusion for the purpose of making better decisions.

Additionally, technology governance at the university level should leverage the culturally and cognitively diverse array of backgrounds, experiences, and competencies that students provide, their distinct technology expertise and lived experience, as well as their specific privacy and data sharing expectations. Specifically, universities can build on the position of students as a distinct political and stakeholder group and institute what we call Student Technology Councils.[242] Our research on student participation in technology decisions across over 6,300 U.S. universities shows that existing student technology councils or similar efforts mirror the approach of engaging students as consumers whose product feedback is solicited by university IT departments, rather than foregrounding deliberation and co-governance. These programs typically form ad hoc groups of student volunteers who are appointed, rather than elected. The Student Technology Council (STC) structure we are proposing, however, would be akin to existing student self-governance bodies, such as Honor Councils, which are responsible for investigating and adjudicating breaches of trust and policies, particularly cases involving serious breaches such as lying, cheating, or stealing. Student self-governance bodies are typically led by student-elected representatives and financially and institutionally supported by the university while remaining independent.

An STC would be comprised of elected student representatives who would not sanction technology use among students, faculty, or administrators, but serve as a representative advising body to faculty and administrators on three key areas: digital technology procurement, digital technology innovation, and digital technology and data governance. These three areas would ensure that student expertise and agency are honored and engaged when considering *what* digital technologies ought to be integrated into university life and infrastructure, *how* to use them to enhance student learning and student life,

242. Mona Sloane, *Biden's AI Executive Order Underlines Need for Student Technology Councils*, TIMES HIGHER EDUC. (Nov. 4, 2023), https://www.timeshighereducation.com/blog/bidens-ai-executive-order-underlines-need-student-technology-councils.

and how to *govern* them so that students' well-being and privacy are prioritized over the imperative of the private industry in technology.

Structurally, the STC would be set up similarly to the German Ethics Council ("Ethikrat"),[243] an independent advisory body mandated in the German federal law and comprised of experts (appointed to four-year terms) that examines ethical, societal, scientific, medical, and legal issues and issues opinions and recommendations to the federal government and parliament on all matters of ethical significance. The Ethics Council convenes monthly and also engages the general public in ethics conversations through public events.[244] Similarly, the STC would be comprised of students who are elected for two-year terms, who convene regularly on technology questions pertaining to the three key areas, and who issue at least one recommendation per semester while also regularly engaging the student body in technology questions on these three key areas (for example, via town halls). Additionally, the STC would advise the university on forming student-inclusive committees that touch upon technology questions on the departmental or university level, such as questions pertaining to AI use policies or data privacy in collegiate athletics. This concrete governance intervention is supported by deliberative decision-making theory and ensures universities make better and more inclusive technology decisions that build on the long history of student self-governance and participation in university governance.

VI. CONCLUSION

The integration of digital technologies, including AI, into university life has exposed deeper, longstanding fractures in the governance structures of higher education—particularly the marginalization of students in decisions that profoundly shape their life inside and outside of the classroom. The interlocking trends of growing pervasive digital technology on campuses and rapidly declining student participation in university governance put students at risk of over-surveillance, profiling, and privacy violations. Similarly, by limiting student governance, universities risk foregoing building on students' knowledge and expertise as early technology adopters and active users, and as experts in their own lives and learning. As digital technologies become ubiquitous and students become increasingly dependent on, and affected by, data-intensive systems that they neither choose nor control, a technology governance pivot is becoming urgent. Decision-making theory and historical precedent both suggest that more inclusive, participatory models of

243. *The Ethics Council*, DEUTSCHER ETHIKRAT [GERMAN ETHICS COUNCIL], https://www.ethikrat.org/en/about-us/the-german-ethics-council/ (last visited Aug. 18, 2025).

244. *Id.*

governance lead to better outcomes—not only for innovation and efficacy, but for legitimacy and trust. Furthermore, there is no legal reason for continuing to exclude students from technology governance. Against that backdrop, establishing Student Technology Councils that are comprised of elected student representatives on two-year terms and that advise faculty and administrators on technology procurement, innovation, and governance can position students as co-governors rather than passive users.