TL;DR: THE LAW AND LINGUISTICS OF SOCIAL PLATFORM TERMS-OF-USE

Tim R Samples,† Katherine Ireland†† & Caroline Kraczon†††

ABSTRACT

Online terms-of-use (TOUs) are the most widely used form contracts in human history. But TOUs are as poorly understood as they are ubiquitous. Their proliferation has fueled a yawning gap between contract law and consumer reality. The notion that users read and understand online TOUs, disproven in academic research, is the subject of pop culture mockery. Yet contract law assumes something very different. Because classic legal doctrines apply to online contracts, consumers routinely find themselves legally bound to contracts they have not—and often could not—read.

In this Article, we evaluate the law and linguistics of a critical area of consumer contracting: smartphone-based social platforms. Our interdisciplinary study examines an original dataset of 195 contracts (TOUs, privacy policies, and community guidelines) for seventy-five apps. Our analysis highlights a decoupling of contract doctrine and consumer reality in the smartphone age of online contracting. Our results show that this divergence is fueled by extraordinary volume, complexity, and asymmetries in platform-to-consumer contracts. In addition, our results offer evidence that the decoupling has grown in recent years.

TABLE OF CONTENTS

I.   INTRODUCTION

Too long; didn't read—often abbreviated as TL;DR or TLDR—is a popular expression indicating that a passage of text appears excessive in length, presumptively not worth reading.[1] For most consumers, routine transactional agreements, like a privacy policy for a smartphone application (or "app"), are the epitome of TL;DR. The terms-of-use (TOUs) of popular social platforms bind billions of users in contracts that govern sensitive personal rights and intimate data.[2] With almost three billion users, Facebook's TOU is among the most widely used contracts ever.[3] In the absence of intervening regulation, TOUs govern much of society's relationship with technology.[4] Yet only a small fraction of users will ever read or understand them.[5] Indeed, the most widespread contracts in the history of the world are among the least understood.[6]

Prominent legal minds—including the Chief Justice of the Supreme Court—have confessed to glossing over the terms of their own consumer contracts.[7] A survey of consumer law scholars indicates a similar pattern.[8]

---

1. The abbreviation is used widely enough to have an entry in Merriam-Webster. *See* TL;DR, MERRIAM-WEBSTER, https://www.merriam-webster.com/dictionary/TL;DR (last visited June 7, 2022).

2. In this Article, we use "terms of use" (abbreviated as TOU) to refer to the variety of standard form agreements that platforms employ in contracting with users. Sometimes these agreements are also called terms of service (e.g., Snapchat, X (formerly Twitter), YouTube), user agreements (e.g., LinkedIn, Helo), license agreements (e.g., OK), and so on.

3. *See Terms of Service*, FACEBOOK, https://www.facebook.com/terms.php (last updated July 26, 2022).

4. *See* Nancy S. Kim & D. A. Jeremy Telman, *Internet Giants as Quasi-Governmental Actors and the Limits of Contractual Consent*, 80 MO. L. REV. 724, 754 (2015); Michael L. Rustad & Thomas H. Koenig, *Wolves of the World Wide Web: Reforming Social Networks' Contracting Practices*, 49 WAKE FOREST L. REV. 1431, 1432–33 (2014).

5. *See* Yannis Bakos, Florencia Marotta-Wurgler & David R. Trossen, *Does Anyone Read the Fine Print? Consumer Attention to Standard Form Contracts*, 43 J. LEGAL STUD. 1, 3 (2014) (demonstrating that consumers almost never read end-user license agreements); *see also* Uri Benoliel & Shmuel I. Becher, *The Duty to Read the Unreadable*, 60 B.C. L. REV. 2255, 2277–81 (2019) (showing that TOUs on popular websites are mostly unreadable for the general public).

6. *See* Michael L. Rustad & Maria Vittoria Onufrio, *Reconceptualizing Consumer Terms of Use for a Globalized Knowledge Economy*, 14 U. PA. J. BUS. L. 1085, 1086 (2012) (labeling social media TOUs as "the most widely used standard form contract in world history, with potentially billions of users").

7. *See* Debra Cassens Weiss, *Chief Justice Roberts Admits He Doesn't Read the Computer Fine Print*, A.B.A. J. (Oct. 20, 2010), https://www.abajournal.com/news/article/chief_justice_roberts_admits_he_doesnt_read_the_computer_fine_print.

8. *See* Jeff Sovern, *The Content of Consumer Law Classes III*, 22 J. CONSUMER & COM. L. 2, 4 ("Not one professor reported always reading contracts or disclosures. In contrast, 57% said they rarely or never read contracts and 48% said they rarely or never read required disclosures.").

Anecdotal evidence is consistent with academic research: consumers almost never read TOUs and, when they do, are unlikely to fully understand their terms.[9] TOUs are, fundamentally, TL;DR. Popular culture ridicules the idea of reading them.[10] A satirical headline in *The Onion* reads, "New Facebook Terms of Service Includes Compulsory Conscription Into Zuckerberg's Upcoming War Against Government."[11] A parody podcast titled Ts&Zzz aims to lull listeners to sleep by reading TOUs aloud in their entirety.[12] Some companies have included amusing clauses or even cash prizes in TOUs as surprises for the rare consumer who actually reads them.[13]

This Article offers an interdisciplinary analysis of the consumer contracting ecosystem, with a focus on social platforms. To start, we construct an original dataset that includes the core consumer contracts of seventy-five digital platforms: their TOUs, privacy policies, and community guidelines. Our dataset contains 195 separate agreements that amount to roughly 944,459 words—almost 4.5 times the length of *Crime and Punishment* by Fyodor Dostoyevsky.[14] Whereas most interdisciplinary work on form contracting has come from law and economics, we use a law and linguistics framework to examine the platform-consumer contracts in our dataset.[15] Our methods combine legal analysis with natural language processing, data science, and corpus linguistics.[16] We supplement traditional readability metrics with more advanced linguistics methods to assess the linguistic difficulty of our dataset.

---

9. *See* Bakos et al., *supra* note 5 and accompanying text.

10. *See*, *e.g.*, *South Park: Human CentiPad*, COMEDY CENTRAL (Apr. 27, 2011), https://www.southparkstudios.com/episodes/j6a6zs/south-park-humancentipad-season-15-ep-1.

11. *New Facebook Terms of Service Includes Compulsory Conscription into Zuckerberg's Upcoming War Against Government*, ONION (Oct. 1, 2019), https://www.theonion.com/new-facebook-terms-of-service-includes-compulsory-inscr-1838675822.

12. Ts&ZZZ, https://tsandzzz.com (last visited Jan. 19, 2024) ("Ts&Zzz is a podcast to help you fall asleep by listening to a conversation about the most boring text on the internet; [sic] terms of service agreements.").

13. *See*, *e.g.*, Matthew S. Schwartz, *When Not Reading the Fine Print Can Cost Your Soul*, NPR (Mar. 8, 2019), https://www.npr.org/2019/03/08/701417140/when-not-reading-the-fine-print-can-cost-your-soul (reporting on various novelty clauses).

14. *See infra* Section III.A (explaining characteristics of the dataset).

15. *See* Shmuel I. Becher, *Asymmetric Information in Consumer Contracts: The Challenge That Is Yet to Be Met*, 45 AM. BUS. L.J. 723, 724 (2008) ("The interdisciplinary contribution to the field of [standard form contracts] is dominated by economics.").

16. *See* Anya Bernstein, *Legal Corpus Linguistics and the Half-Empirical Attitude*, 106 CORNELL L. REV. 1397, 1401–17 (2021) (describing the emergence and development of corpus linguistics research in fields of linguistics and law). Our methods rely on R programming packages including quanteda, tidytext, and polmineR. *See* Section III.B (explaining our methods).

We also analyze metadata variables to inform our observations about the consumer contracting practices of social platforms.[17]

This Article also offers a novel contribution to the field of legal corpus linguistics—a promising [18] yet controversial [19] area of empirical legal scholarship. Corpus linguistics is the scientific study of naturally occurring language in the aggregate, often in large datasets, so-called corpora.[20] It applies a variety of computational and quantitative methodologies to understand fine-grained and large-scale trends across different linguistic levels of analysis. Previous work in legal corpus linguistics has focused primarily on matters of judicial interpretation, such as the ordinary meaning of specific words.[21] We take a different tack, combining legal analysis with methods from corpus linguistics, data science, and natural language processing. Our interdisciplinary methodology contributes to a vibrant and growing literature that evaluates online consumer contracts with empirical methods.[22]

Our methods bridge a gap between law and linguistics. Legal scholarship has produced a wealth of literature on the law and problems of consumer contracting. Linguistics scholarship, meanwhile, offers a wealth of advanced

---

17. *See infra* Section IV.D (discussing the results of our metadata analysis).

18. *See*, *e.g.*, Friedemann Vogel, Hanjo Hamann & Isabelle Gauer, *Computer-Assisted Legal Linguistics: Corpus Analysis as a New Tool for Legal Studies*, 43 LAW & SOC. INQUIRY 1340, 1340 (2018) (cataloging research on corpus linguistics and introducing "computer-assisted legal linguistics"); Thomas R. Lee & Stephen C. Mouritsen, *Judging Ordinary Meaning*, 127 YALE L.J. 788, 795 (2018) (proposing the use of corpus linguistic methods to ascertain the ordinary meaning of texts for judicial interpretation); James C. Phillips, Daniel M. Ortner & Thomas R. Lee, *Corpus Linguistics & Original Public Meaning: A New Tool to Make Originalism More Empirical*, 126 YALE L.J. F. 21, 21 (2016) (same).

19. *See*, *e.g.*, Bernstein, *supra* note 16, at 1399 (highlighting weaknesses in legal corpus linguistics research that "hindered its own ability to yield empirically reliable results"); John S. Ehrett, *Against Corpus Linguistics*, 108 GEO. L.J. ONLINE 50 (2019) (arguing against the use of corpus linguistics in judicial interpretation); Evan C. Zoldan, *Corpus Linguistics and the Dream of Objectivity*, 50 SETON HALL L. REV. 401, 401 (2019) (concluding that "corpus linguistics does not live up to its promise to make legal interpretation more objective").

20. *See* Bernstein, *supra* note 16, at 1402–12; Lee & Mouritsen, *supra* note 18, at 795.

21. *See*, *e.g.*, Carpenter v. United States, 138 U.S. 2206, 2238 (2018) (Thomas, J., dissenting) (citing legal corpus linguistics research on the meaning of "search"); Am. Bankers Ass'n v. Nat'l Credit Union Admin., 306 F.3d 44, 68 n.5 (D.C. Cir. 2018), *rev'd and remanded*, 934 F.3d 649 (D.C. Cir. 2019) (citing to the Corpus of Historical American English on the meaning of "rural district"); *see also* Zoldan, *supra* note 19, at 404–05 (reviewing instances of judges using and citing corpus linguistics methods in judicial opinions).

22. *See*, *e.g.*, Shmuel I. Becher & Uri Benoliel, *Dark Contracts*, 64 B.C. L. REV. 55, 69 n.78 (2023) (reviewing empirical research on form contracts); Rustad & Koenig, *supra* note 4 (evaluating the contracting practices of social platforms); *see* Florencia Marotta-Wurgler & Robert Taylor, *Set in Stone? Change and Innovation in Consumer Standard-Form Contracts*, 88 N.Y.U. L. REV. 240 (2013) (examining change in consumer form contracts).

metrics for understanding the linguistic difficulty of texts.[23] To date, advanced methods in linguistics have only begun to reckon with issues in consumer contracting.[24] In this Article we take a step in that direction. Our analysis examines three categories of linguistic characteristics that may contribute to reading difficulty: *readability* (sentence and word length), *syntactic complexity* (range and complexity of language forms), and *lexical diversity* (richness of vocabulary).[25]

Contracting has changed profoundly during the digital era, but contract law has not. For courts, online TOUs have proven particularly awkward to evaluate through lenses of traditional contract law.[26] As an extension of internet contracting, the smartphone generation of platform-consumer TOUs has introduced new distortions to traditional contract doctrines. Our results illustrate that trend. In this Article, we demonstrate the extraordinary volume, linguistic difficulties, and asymmetries facing consumers online. By situating our results in longitudinal comparisons with similar datasets, we also show that these tendencies have deepened in recent years.[27] All said, the sum of our findings supports a broader conclusion: the gap between contract doctrine and consumer reality—already wide in the online environment—has grown wider

---

23. Readability metrics emerged in the 1940s to measure the ease of reading a text in quantitative terms. Since then, however, linguistics scholarship has yielded a more diverse set of metrics for understanding the linguistic difficulty of a text. *See generally* Matej Martinc, Senja Pollak & Marko Robnik-Šikonja, *Supervised and Unsupervised Neural Approaches to Text Readability*, 47 COMPUTATIONAL LINGUISTICS J. 141 (2021); Tove Larsson & Henrik Kaatari, *Syntactic Complexity Across Registers: Investigating (in)formality in Second-Language Writing*, 45 J. ENGLISH FOR ACADEMIC PURPOSES 1 (2020); Mostafa Zamanian & Pooneh Heydari, *Readability of Texts: State of the Art*, THEORY AND PRAC. IN LANGUAGE STUD. (2012); Michael A. Covington & Joe D. McFall, *Cutting the Gordian Knot: The Moving-Average Type-Token Ratio (MATTR)*, 17 J. QUANTITATIVE LINGUISTICS 94 (2010); Eric Martínez, Francis Mollica & Edward Gibson, *Poor Writing, Not Specialized Concepts, Drives Processing Difficulty in Legal Language*, 224 COGNITION 1 (2022).

24. *See generally* Martínez et al., *supra* note 23 (examining features of contract language that inhibit processing and comprehension); Isabel Wagner, *Privacy Policies Across the Ages: Content and Readability of Privacy Policies 1996–2021*, https://doi.org/10.48550/arXiv.2201.08739 (2022) (using longitudinal data to assess linguistic tendencies and other features of privacy policies).

25. *See infra* Section III.B (defining and illustrating our metrics).

26. *See*, *e.g.*, Berkson v. Gogo LLC, 97 F. Supp. 3d 359, 380 (E.D.N.Y. 2015) ("Courts have 'decided,' based largely on speculation, what constitutes inquiry notice of a website's 'terms of use.'").

27. *See infra* Sections IV.A–D (comparing our results with previous studies of online TOUs).

in the smartphone era.[28] That divergence is especially problematic for social platforms that transact with society at extraordinary scale, deploy manipulative interfaces, and manage vast troves of intimate user data.[29]

This Article proceeds as follows. Part II describes platform TOUs and their place within the modern consumer contracting landscape. Following that overview, we outline features of social platforms and smartphones that compound the implications of these TOUs for individual rights and public interests. In this way, we differentiate mobile-based social platform TOUs from other areas of consumer contracting. In Part III, we explain our data and methodology. We begin by explaining the characteristics of our dataset and our approach to gathering the data. We then discuss the metrics and computations in our methodology. Part IV presents the findings. There, we illustrate and discuss the results of our linguistics analysis. We also use our metadata to demonstrate key tendencies among social platform TOUs. As we discuss the results, we consider implications for law and policy. A brief conclusion follows.

II.   TOUS AND SOCIAL PLATFORMS

Standard form contracting is now the primary means for conducting business in consumer-facing industries.[30] With the Industrial Revolution, mass production and distribution prompted the need for standardized contracts.[31] By reducing transaction costs, form contracting at scale offers important efficiency gains. But many of the benefits of form contracting are deeply asymmetric, producing tensions with fundamental tenets of contract law.[32] The nature of online contracting combined with the unprecedented scale of

---

28. *See* Jeff Sovern & Nahal Heydari, *Not-So-Smartphone Disclosures*, St. John's Legal Studies Research Paper No. 22-0010 (Aug. 12, 2022), http://dx.doi.org/10.2139/ ssrn.4188892.

29. *See infra* Section II.B (distinguishing social platform contracting from other areas of consumer contracting).

30. Daniel T. Ostas, *Postmodern Economic Analysis of Law: Extending the Pragmatic Visions of Richard A. Posner*, 36 AM. BUS. L.J. 193, 226–27 (1998) ("Form contracting has become the predominant way of doing business in the twentieth century.").

31. Becher, *supra* note 15, at 726; Ellen Wauters, Eva Lievens & Peggy Valcke, *Towards a Better Protection of Social Media Users: A Legal Perspective on the Terms of Use of Social Networking Sites*, 22 INT'L J. L. INFO. TECH. 254, 255 (2014).

32. Legal scholars have critiqued standard form consumer contracts for decades. *See generally* Friedrich Kessler, *Contracts of Adhesion—Some Thoughts about Freedom of Contract*, 43 COLUM. L. REV. 629 (1943); W. David Slawson, *Standard Form Contracts and Democratic Control of Lawmaking Power*, 84 HARV. L. REV. 529 (1971); Melvin A. Eisenberg, *The Limits of Cognition and the Limits of Contracts*, 47 STAN. L. REV. 211 (1995); Robert A. Hillman, *Rolling Contracts*, 71 FORDHAM L. REV. 743 (2002).

digital platforms further exacerbate those tensions. This Part provides a brief overview of the legal environment for consumer contracting online. Following that overview, this Part outlines characteristics that differentiate the TOUs of social platforms from other areas of consumer contracting.

A.   TOUS IN THE LEGAL ENVIRONMENT

This Section II.A discusses both the development of modern TOUs and TOU typology within legal environments.

    1.   *Modern TOUs*

Form contracting has long been controversial.[33] Yet, for today's consumer, form contracts are more pervasive than ever before. Browsing websites, making routine purchases, downloading an app—virtually any online activity involves a TOU, a privacy policy, or both.[34] Online commerce and mobile computing created vast new frontiers for consumer contracting.[35] The debut of the iPhone in 2007 gave rise to a new era of digital commerce on mobile devices. When the App Store launched in 2008, it carried about 500 apps. Today, the App Store offers almost 2.2 million apps while Google Play has over 3.5 million.[36] For people across the world, information access underwent a profound shift towards mobile, app-based web experiences.[37]

As the app economy grew into a multi-trillion-dollar marketplace, a vast and ever-expanding universe of consumer contracts followed.[38] Digital

---

33. For a colorful critique from the 1940s, *see* Kessler, *supra* note 32, at 640 ("Standard contracts in particular could thus become effective instruments in the hands of powerful industrial and commercial overlords enabling them to impose a new feudal order of their own making upon a vast host of vassals.").

34. Ann Morales Olazábal, Howard Marmorstein & Dan Sarel, *Frequent Flyer Programs: Empirically Assessing Consumers' Reasonable Expectations*, 51 AM. BUS. L.J. 175, 221 (2014) ("As early as the 1970s, standardized contract forms had already edged out the practice of contract negotiation, with the vast majority of consumer and commercial contracts being form-driven.").

35. Woodrow Hartzog, *Website Design as Contract*, 60 AM. U. L. REV. 1635, 1641 (2011) ("As websites became ubiquitous, so did terms of use. As a result, an overwhelming amount of online activity is not governed by default law but rather through agreement between the parties.").

36. The Amazon Appstore has another 480,000 or so. *See* L. Ceci, *Number of Apps Available in Leading App Stores as of 2nd Quarter 2022*, STATISTA (Aug. 11, 2022), https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores.

37. *See* ORG. FOR ECON. COOP. & DEV. [OECD], *The App Economy*, at 8–11, OECD Digital Econ. Papers No. 230 (Dec. 16, 2013), http://dx.doi.org/10.1787/5k3ttftlv95k-en (summarizing the rapid growth of the app economy).

38. In 2021, the app economy was worth some $6.3 trillion. *See* L. Ceci, *Size of the App Economy Worldwide from 2016-2021*, STATISTA (July 6, 2021), https://www.statista.com/statistics/267209/global-app-economy.

platforms now cater to millions, or even billions, of consumers at once. The number of active users on the Meta family of platforms—including Facebook, WhatsApp, and Instagram—is almost 3.6 billion per month.[39] According to Pew surveys, a quarter of Americans report that they are asked to agree to a privacy policy daily.[40] Few read them; even fewer read them all the way through.[41]

From a consumer perspective, the online contracting environment is especially daunting. The length of TOUs introduces a fundamental asymmetry between platforms and users. While reading TOUs is costly for consumers, adding terms to an online contract costs a platform almost nothing.[42] There are no physical constraints on the length of an online contract. Over time, TOUs have swelled in length and complexity. Smartphone-based apps pose serious difficulties for reading and comprehension.[43] Practically speaking, to review the user terms of any given platform would mean reading long, highly technical documents on a smartphone screen upon downloading an app.

With the rise of digital commerce, consumer contracting practices have evolved dramatically. Contract law, however, has not.[44] As a result, core concepts in contract law exist in tension with the practical realities of modern form contracting.[45] With a mobile device and an internet connection, contracts may be formed any time, from almost anywhere. However, because the law of contract formation remains relatively static, courts are equipped with outdated

---

39. *Meta Reaches 3.6 Billion People Each Month*, STATISTA (Oct. 29, 2021), https://www.statista.com/chart/2183/facebooks-mobile-users/.

40. *Americans' Attitudes and Experiences with Privacy Policies and Laws*, PEW RSCH. CTR. (Nov. 15, 2019), https://www.pewresearch.org/internet/2019/11/15/americans-attitudes-and-experiences-with-privacy-policies-and-laws.

41. *Id.*

42. *See* Adam Levitin, *ALI Consumer Contracts Restatement—What's at Stake*, CREDIT SLIPS (Mar. 16, 2019), https://www.creditslips.org/creditslips/2019/05/ali-consumer-contracts-restatement-whats-at-stake.html ("[T]he cost of larding on an extra term on the Internet is so low, that there's no reason for a business not to bury its whole Christmas wishlist in linked on-line terms and conditions.").

43. *See infra* Section IV.A–D.

44. *See, e.g.*, Plazza v. Airbnb, Inc., 289 F. Supp. 3d 537, 547 (S.D.N.Y. 2018) ("Although the Internet age has certainly introduced new twists with regard to entering into contracts, the fundamental elements of contract law, including mutual assent of the parties, have not changed."); Register.com, Inc. v. Verio Inc., 356 F.3d 393, 403 (2d Cir. 2004) ("While new commerce on the Internet has exposed courts to many new situations, it has not fundamentally changed the principles of contract.").

45. *See* Becher, *supra* note 15, at 724 ("Because typical consumers do not read and cannot negotiate [standard form contracts], such contracts challenge the basic assumption of informed consent as a prerequisite for contract formation.").

doctrines to the online contracting environment.[46] For instance, case law shaping the doctrine of inquiry notice developed in vending machine disputes in the 1950s and 1960s.[47] Yet inquiry notice remains a pivotal issue in the enforceability of online TOUs.[48]

Most users, of course, do not read TOUs.[49] And few would understand them even if they did.[50] Nonetheless, courts often treat TOUs as valid, enforceable contracts.[51] Whether the consumer has read a TOU is irrelevant.[52] Whether or not the consumer *reasonably could* read a TOU is also irrelevant. Because courts approach modern TOUs with traditional contract doctrines, concepts like reasonable notice are adapted to the online contracting environment.[53] Determinations of notice often turn on small details like font and color scheme, the conspicuousness of hyperlinks, and interface design.[54] Courts tend not to question the practicality or the reasonable feasibility of reading a TOU. The existence of an *opportunity* to read will suffice if notice of the terms is deemed conspicuous.[55] Although the law imposes a duty to read

---

46. *See* NANCY S. KIM, WRAP CONTRACTS: FOUNDATIONS AND RAMIFICATIONS 109–11 (2013) (comparing internet contract rules to traditional contract doctrine); *Berkson*, 97 F. Supp. 3d 359 at 381–82 (observing "outdated fundamentals" available to courts).

47. *See Berkson*, 97 F. Supp. 3d at 382 (explaining the foundations of inquiry notice case law); *see also* Specht v. Netscape, 306 F.3d 17, 30–32, 35 (2d Cir. 2002) (applying a traditional reasonable communicativeness test to an internet contract dispute).

48. *See, e.g.*, Selden v. Airbnb, Inc., 4 F.4th 148, 157 (D.C. Cir. 2021).

49. *See generally* Jonathan A. Obar & Anne Oeldorf-Hirsch, *The Biggest Lie On the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services*, 23 INFO., COMM'N & SOC'Y 128 (2020); Shmuel I. Becher & Tal Z. Zarsky, *Online Consumer Contracts: No One Reads, but Does Anyone Care?*, 12 JERUSALEM REV. LEGAL STUDIES 105 (2010); Susan E. Gindin, *Nobody Reads Your Privacy Policy or Online Contract? Lessons Learned and Questions Raised by the FTC's Action Against Sears*, 8 NW. J. TECH. & INTELL. PROP. 1 (2009); Ian Ayres & Alan Schwartz, *The No-Reading Problem in Consumer Contract Law*, 66 STAN. L. REV. 545 (2014); Bakos et al., *supra* note 5.

50. *See* Bakos et al., *supra* note 5 and accompanying text.

51. When notice and assent are considered adequate, TOUs have widely been ruled enforceable. *See* Selden v. Airbnb, Inc., 4 F.4th 148, 157 (D.C. Cir. 2021).

52. The duty to read doctrine is well established in common law. Even if a party does not read a contract, courts presume that all parties have read it and are bound to its terms, as long as users have adequate notice of the terms and express their assent. *See* Becher, *supra* note 15, at 729–33.

53. *See, e.g.*, Selden v. Airbnb, Inc., 4 F.4th 148, 157–58 (D.C. Cir. 2021).

54. *See, e.g., id.* ("The only red text in the warning indicated the legal policies, which were set off from the surrounding black text."); Meyer v. Uber, 868 F.3d 66, 78 (2d Cir. 2017) ("Turning to the interface at issue in this case, we conclude that the design of the screen and language used render the notice provided reasonable as a matter of California law."); Nicosia v. Amazon.com, Inc., 834 F.3d 220, 233 (2d Cir. 2015).

55. *See, e.g.*, Swift v. Zynga Game Network, Inc., 805 F. Supp. 2d 904, 911–12 (N.D. Cal. 2011).

contracts on consumers, there is no symmetrical duty to make contracts readable or understandable.[56] Gaps like these create tensions for courts, especially in consumer contracting disputes.

Social platform TOUs highlight those tensions. They are deeply asymmetric, procedurally and substantively.[57] Most are contracts of adhesion, offered on a unilaterally drafted, take-it-or-leave-it basis.[58] Many are extraordinary in length and linguistic difficulty.[59] The fact that virtually none of the millions (or billions) of people who agree to a given TOU will ever read or understand the terms raises fundamental legal questions: What constitutes meaningful consent, or even notice, in these conditions? Is this freedom of contract? If so, for who? The idea that TOUs represent a "meeting of the minds" borders on the absurd.[60] Because classic pillars of contract law are so distorted in the online contracting environment, courts are forced into the realm of speculation around central issues.[61]

In theory, the notice and choice framework should empower consumers to make choices about how their personal data will be handled online. But that utopia is far from the reality for consumer TOUs. An uneasy assumption undergirds the enforceability of online TOUs: that consumers receive adequate notice of terms and make free choices about whether to agree with the terms. That assumption is especially precarious in the modern contracting environment where consumers face cognitive hurdles, time constraints, and informational asymmetries that prevent them from making rational choices about whether to agree to TOUs.[62] The volume *and* content of TOUs is overwhelming.[63] And, even if consumers could both read and understand

---

56. *See* Becher, *supra* note 5, at 2258 ("Yet under U.S. law, the duty to read is unilateral: although consumers are presumed to read contracts, there is no general duty on suppliers to provide consumers with readable contracts.").

57. There is a wide gap across a variety of factors: opportunities to draft and negotiate, bargaining power, legal expertise, business sophistication, and commercial experience. *See* Schmuel I. Becher & Esther Unger-Aviram, *The Law of Standard Form Contracts: Misguided Intuitions and Suggestions for Reconstruction*, 8 DEPAUL BUS. & COMM. L.J. 199, 201 (2010); *see also* Becher, *supra* note 15, at 727.

58. *See* Berkson v. Gogo LLC, 97 F. Supp. 3d 359, 365 ("In many instances, these consumers are accepting important contracts of adhesion when they order a product or service through a computer.").

59. *See infra* Sections IV.A–D (illustrating the difficulty of TOUs relative to other bodies of English usage and their expanding length).

60. Rustad & Koenig, *supra* note 4, at 1434 ("The concept of the 'meeting of the minds' is a legal fiction when it comes to TOU boilerplate.").

61. *See, e.g.*, Berkson v. Gogo LLC, 97 F. Supp. 3d 359, 380.

62. *See* Daniel J. Solove, *Introduction: Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880, 1883 (2013).

63. *See infra* Part IV.

TOUs, they still cannot negotiate their terms.[64] Notice and assent are assumed via legal constructs but are largely a fantasy in today's commercial and technological environment.

2. TOU Typology

Social platform TOUs belong to the family of "wrap" contracts.[65] Nancy Kim defines the family as a "unilaterally imposed set of terms which the drafter purports to be legally binding."[66] Wrap contracts are presented in non-traditional formats—a signature is not required, nor is a pen. There are three classic wrap forms: shrink-wraps, click-wraps, and browse-wraps.[67] There are also scroll-wraps, sign-in-wraps, and various other creatures in the wrap family.[68] As a matter of terminology, the use of the word "wrap" is a relic from an earlier era of consumer contracting practices. The origin lies with shrink-wrap contracts, which derive their name from the cellophane packaging on software product boxes.[69] Manufacturers often included a license agreement on the box, visible through the cellophane wrapper.[70] In that form, the consumer accepts the contract by breaking the seal to open the box.[71] For whatever reason, the word "wrap" has persisted even as non-traditional contracting practices have evolved into digital forms.

Social platform TOUs are often considered sign-in-wrap agreements, which combine features of click-wraps with browse-wraps.[72] Sign-in-wraps present a digital prompt that indicates agreement with an online contract, which is often hyperlinked on an account registration panel.[73] Usually, a sign-in-wrap provides that, by signing up for an account, the user agrees to the contract (and other supplementary terms, such as privacy policies). Click-

---

64. Solove, *supra* note 62, at 1888.

65. *See* KIM, *supra* note 46, at 2.

66. *Id.*

67. *Id.* at 36–43 (providing a typology of the classic wrap forms).

68. Click-wraps, for instance, may also be called "click-through," "click and accept," or "web-wrap" agreements.

69. *The Origin of Click-Wrap: Software Shrink-Wrap Agreements*, WILMER HALE (Mar. 22, 2000), https://www.wilmerhale.com/insights/publications/the-origin-of-click-wrap-software-shrink-wrap-agreements-march-22-2000.

70. *Id.*

71. After a number of cases finding shrink-wrap agreements unenforceable, a seminal case in the Seventh Circuit established their enforceability—and a circuit split on the question. *See* ProCD, Inc. v. Zeidenberg, 86 F.3d 1447 (7th Cir. 1996); *see also* Brian Covotta & Pamela Sergeeff, ProCD, Inc. v. Zeidenberg, 13 BERKELEY TECH. L.J. 35, 36–37 (1998) (reviewing early case law on the enforceability of shrink-wrap contracts).

72. Selden v. Airbnb, Inc., 4 F.4th 148, 156–57 (D.C. Cir. 2021) ("[A] sign-in wrap bundles signing up for a service with agreement to the website's contractual terms.").

73. Berkson v. Gogo LLC, 97 F. Supp. 3d 359, 399–401 (E.D.N.Y. 2015).

wraps, meanwhile, prompt a user to perform a digitally-mediated action, such as tapping or clicking an "I agree" button, that indicates assent.[74] Browse-wraps, on the other hand, are more passive. A typical browse-wrap is a statement (hyperlinking to the actual TOU) at the bottom of a screen that says using the website amounts to acceptance. Importantly, browse-wraps do not require a proactive confirmation of assent.[75] In theory, at least, consumers have an opportunity to review click-wrap and sign-in-wrap agreements before using the platform or service. In court, form matters: whereas judges have shown reluctance to enforce browse-wrap agreements, click-wraps and sign-in-wraps tend to be more enforceable.[76]

B.   SOCIAL PLATFORM TOUs

Social platform TOUs are a distinct and particularly sensitive area of consumer contracting. As take-it-or-leave-it agreements, platform TOUs resemble, in many ways, routine form contracts in the digital era. But social platform TOUs carry distinct implications—above and beyond longstanding dilemmas posed by form consumer contracting more broadly. We develop four points within that proposition, highlighting characteristics particular to this category of TOUs: (1) social networks operate and contract at an unprecedented, systemic scale;[77] (2) social platform business models often rely

---

74. Jonathan A. Obar & Anne Oeldorf-Hirsch, *The Clickwrap: A Political Economic Mechanism for Manufacturing Consent on Social Media*, SOCIAL MEDIA & SOC'Y 1, 3 (2018) ("The clickwrap is a digital prompt that enables the user to provide or withhold their consent to a policy or set of policies by clicking a button, checking a box, or completing some other digitally mediated action suggesting 'I agree' or 'I don't agree'"); Eric Goldman, *How Zappos' User Agreement Failed In Court and Left Zappos Legally Naked*, FORBES (Oct. 10, 2012), https://www.forbes.com/sites/ericgoldman/2012/10/10/how-zappos-user-agreement-failed-in-court-and-left-zappos-legally-naked ("A clickthrough agreement is presented to users in such a way that they must take some action—usually, clicking on a button—that unambiguously signifies that they are assenting to the contract.").

75. KIM, *supra* note 46, at 41 ("Browsewraps do not require users to affirmatively manifest consent.").

76. *See, e.g.*, Berman v. Freedom Financial Network, LLC, 30 F.4th 849, 856 (9th Cir. 2022) ("Courts are more reluctant to enforce browsewrap agreements because consumers are frequently left unaware that contractual terms were even offered, much less that continued use of the website will be deemed to manifest acceptance of those terms."); Sellers v. JustAnswer LLC, 73 Cal. App. 5th 444, 466 (2021) ("[Courts] have reached consistent conclusions when evaluating the enforceability of agreements at either end of the spectrum, generally finding scrollwrap and clickwrap agreements to be enforceable and browsewrap agreements to be unenforceable.").

77. *See, e.g.*, Nizan Geslevich Packin, *Too-Big-to-Fail 2.0? Digital Service Providers as Cyber-Social Systems*, 93 INDIANA L.J. 1211, 1216 (2018) (exploring failure risks among "Critical Service Providers" in digital markets); Carl Öhman & Nikita Aggarwal, *What if Facebook Goes Down? Ethical and Legal Considerations for the Demise of Big Tech*, 9 INTERNET POL'Y REV. 1, 6 (2020) (coining the term "Systemically Important Technology Institution" and calling for

on harvesting consumer attention with manipulative designs;[78] (3) heavy market concentrations limit and bind consumer choices in the social network marketplace; and (4) in legal systems with weak consumer and data protections—the United States, for instance—social platform TOUs have an outsized role in defining society's relationship with technology.[79]

1. *Unprecedented Scale*

The scale of social platform contracting is both extraordinary and unprecedented. Previous eras of form contracting appear quaint by comparison. The largest platforms contract with consumers by the billions, compounding the implications of their TOUs. Facebook's TOU alone applies to almost three billion users—equivalent to well over a third of the world's population.[80] YouTube's terms apply to well over two billion users and X's (formerly Twitter) TOU covers some 290 million accounts. Even the TOUs of relatively niche platforms can have vast reach: Badoo (318 million), Tinder (seventy-five million), and Venmo (seventy million). Contracting at this volume has systemic implications.[81] Platform TOUs create private law at societal scale, binding billions of people in contracts that govern sensitive user data and human rights.

2. *Attention-Surveillance Business Models*

The data collection capabilities of digital platforms raise a wide variety of dilemmas, from privacy rights to national security. A common denominator

---

deeper consideration of the concept in digital markets); Lindsay Sain Jones & Tim R Samples, *On the Systemic Importance of Digital Platforms*, 25 U. PA. J. OF BUS. L. 141, 148–49 (2023) (reviewing literature on the concept of systemic importance in the digital realm).

78. *See, e.g.*, Caleb N. Griffin, *Systemically Important Platforms*, 107 CORNELL L. REV. 445, 449–50 (2022).

79. *See* Kim & Telman, *supra* note 4, at 754 ("The business practices of Internet giants set online standards, restrict or delete consumers' rights, establish business norms, and dictate behavior that shapes and affects the lives of citizens."); Jones & Samples, *supra* note 77, at 170 ("Though seemingly mundane, TOUs play a large role in defining legal dynamics—including rights to data, dispute resolution, and privacy—between society and technology.").

80. S. Dixon, *Number of Monthly Active Facebook Users Worldwide as of 2nd Quarter 2022*, STATISTA (July 28, 2022), https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide ("With roughly 2.93 billion monthly active users as of the second quarter of 2022, Facebook is the most used online social network worldwide."). As another point of reference Meta's active *monthly* user base is roughly equivalent to the combined inhabitants of the five largest countries by population: China, India, United States, Indonesia, and Pakistan. Meta reports 3.65 billion monthly active users among its "family" of social applications. Meta's "family" of apps is defined as Facebook, Instagram, Facebook Messenger, and WhatsApp; *see* Felix Richter, *Meta Reaches 3.6 Billion People Each Month*, STATISTA (Oct. 29, 2021), https://www.statista.com/chart/2183/facebooks-mobile-users.

81. *See supra* note 77 and accompanying text.

across those dilemmas is the attention-surveillance business model.[82] A primary aim of that business model is to extract data from users while capturing their attention.[83] Advertising revenues fuel the business: specifically, targeted ads are sold to third parties. Meta, for instance, is almost exclusively reliant on ad sales.[84]

The attention industry predates the digital era by a large margin, with roots in nineteenth-century marketing, propaganda, and ad-based media.[85] Accelerated by the release of the iPhone in 2007, the mobile computing revolution brought powerful technology much closer to the human experience.[86] By 2015, two-thirds of Americans owned a smartphone; by 2021, about 85% did.[87] As humans began living in close proximity to computers around the clock, opportunities for data collection flourished.[88] Put one way, carrying around "the most sophisticated tracking and monitoring device ever forged by the hand of man" has serious implications for privacy.[89] The

---

82. Julie Cohen, *Law for the Platform Economy*, 51 U.C. DAVIS. L. REV. 133 (2017); Shoshanna Zuboff, THE AGE OF SURVEILLANCE CAPITALISM: THE RIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER (2018).

83. Tristan Harris, *Big Tech's Attention Economy Can Be Reformed. Here's how.*, MIT TECH. REV. (Jan. 10, 2021), https://www.technologyreview.com/2021/01/10/1015934/facebook-twitter-youtube-big-tech-attention-economy-reform ("News feeds on Facebook or Twitter operate on a business model of commodifying the attention of billions of people per day"). Platforms that rely on subscription revenues have a different incentive structure. Sara Brown, *The Case for New Social Media Business Models*, IDEAS MADE TO MATTER (June 16, 2021), https://mitsloan.mit.edu/ideas-made-to-matter/case-new-social-media-business-models.

84. In 2020, approximately 97.9% of Facebook's global revenue came from advertising. Stacy Jo Dixon, *Meta's (formerly Facebook Inc.) advertising revenue worldwide from 2009 to 2021*, STATISTA (July 27, 2022), https://www.statista.com/statistics/271258/facebooks-advertising-revenue-worldwide.

85. *See* TIM WU, THE ATTENTION MERCHANTS: THE EPIC SCRAMBLE TO GET INSIDE OUR HEADS 11–23 (2016) (outlining the early eras of attention industries).

86. Gabriella M. Harari, Nicholas D. Lane, Rui Wang, Benjamin S. Crosier, Andrew T. Campbell & Samuel D. Gosling, *Using Smartphones to Collect Behavioral Data in Psychological Science: Opportunities, Practical Considerations, and Challenges*, 11 PERSPECTIVES PSYCH. SCI. 838, 838–39 (2016) ("[Smartphones] are sensor-rich, computationally powerful, and nearly constant companions to their owners, providing unparalleled access to people as they go about their daily lives.").

87. Aaron Smith, *U.S. Smartphone Use in 2015*, PEW RSCH. CTR. (Apr. 1, 2015) https://www.pewresearch.org/internet/2015/04/01/us-smartphone-use-in-2015; *Percentage of U.S. Adults Who Own a Smartphone 2011-2022*, STATISTA, https://www.statista.com/statistics/219865/percentage-of-us-adults-who-own-a-smartphone (last visited Nov. 13, 2023).

88. *See, e.g.*, Harari et al., *supra* note 86 and accompanying text.

89. Alex Kingsbury, *We're About to Find Out What Happens When Privacy Is All but Gone*, N.Y. TIMES (Aug. 23, 2022), https://www.nytimes.com/2022/08/23/opinion/apple-internet-privacy-tracking.html.

"internet of bodies" and the "internet of things" enable the harvesting of even more intimate information, including sensitive health and biometric data.[90]

Mobile-based apps are particularly powerful, combining the intimate user data available on social platforms with the surveillance capabilities of smartphones. Smartphone apps frequently request intimate information such as location tracking, camera access, purchase history, financial information, SMS messages, contacts, various forms of user content, phone call logs, and so on. [91] Geolocation alone offers tremendous opportunities for data collection. In *Carpenter v. United States*, the Supreme Court discussed privacy risks of location tracking, explaining that it can reveal "not only [one's] particular movements, but through them [one's] 'familial, political, professional, religious, and sexual associations.'"[92]

A recent action by the Federal Trade Commission (FTC) against Kochava Inc., an ad tech company based in Idaho, provides further illustration.[93] The FTC complaint outlined sensitivities associated with geolocation data, such as the ability to connect users with locations related to "medical care, reproductive health, religious worship, mental health, temporary shelters, such as shelters for the homeless, domestic violence survivors, or other at-risk populations, and addiction recovery."[94] Kochava's ad tech business model, of course, is far from unique in its use of geolocation data.[95]

Social platforms are positioned to harvest particularly intimate and sensitive data about their users. Facebook might have the "broadest, deepest, and most comprehensive" dataset of human information ever assembled.[96] At times, our apps know us better than we know ourselves and each other. Platforms that engage in behavioral targeting develop extensive profiles on

---

90. Stacy-Ann Elvy, *Commodifying Consumer Data in the Era of the Internet of Things*, 59 B.C. L. REV. 423, 426–27 (2018) (explaining that such devices can collect data "such as fingerprint scans, facial scans, heart rates, fitness levels, temperature, and blood sugar levels, among other things").

91. *See* Gillian Cleary, *Mobile Privacy: What Do Your Apps Know About You?*, SYMANTEC ENTER. BLOGS: THREAT INTEL. (Aug. 16, 2018), https://symantec-enterprise-blogs.security.com/blogs/threat-intelligence/mobile-privacy-apps.

92. Carpenter v. United States, 138 U.S. 2206, 2217 (2018) (quoting United States v. Jones, 565 U.S. 400 (2011) (Sotomayor, J. concurring)).

93. Complaint for Permanent Injunction and Other Relief, F.T.C. v. Kochava Inc., No. 2:22-cv-377, https://www.ftc.gov/system/files/ftc_gov/pdf/1.%20Complaint.pdf.

94. *Id.* at 1–2.

95. David Shepardson, *U.S. Agency to Probe How Mobile Carriers Use Consumer Location Data*, REUTERS (Aug. 25, 2022), https://www.reuters.com/business/media-telecom/us-fcc-investigate-mobile-carrier-use-consumer-geolocation-data-2022-08-25.

96. Jon Evans, *When Facebook Knows You Better Than You Know Yourself*, TECHCRUNCH (Oct. 25, 2015), https://techcrunch.com/2015/10/24/when-facebook-knows-you-better-than-you-know-yourself.

users based on tastes, preferences, and personalities.[97] Data-generated "character" scores can assess credit risks and personality tendencies.[98] The amount of consumer data transferred in a routine transaction—take, for instance, pizza and a movie at home—is staggering.[99] TikTok, as explained by the head of the Federal Communications Commission (FCC), gathers an extremely rich set of user data:[100]

> Indeed, TikTok collects everything from search and browsing histories to keystroke patterns and biometric identifiers, including faceprints—which researchers have said might be used in unrelated facial recognition technology—and voiceprints. It collects location data as well as draft messages and metadata, plus it has collected the text, images, and videos that are stored on a device's clipboard.

Dating platforms also gather troves of sensitive information about their users: full name, age, email address, credit card, geolocation, user photos and videos, political views, religious beliefs, employment and education, social media, chat history, swiping records, behavioral data, marital status, ethnicity, hobbies and interests, gender, sex, sexual orientation, and mobile number/device.[101] Some apps even collect height, weight, and HIV status.[102] And the sheer volume of data harvesting is staggering. One journalist who

---

97. *See* Jared S. Livingston, *Invasion Contracts: The Privacy Implications of Terms of Use Agreements in the Online Social Media Setting*, 21 ALB. L.J. SCI. & TECH. 591, 596 (2011) (explaining the practice of behavioral targeting).

98. *See, e.g.*, Janine S. Hiller & Lindsay Sain Jones, *Who's Keeping Score: Oversight of Changing Consumer Credit Infrastructure*, 59 AM. BUS. L.J. 61, 62–65 (2022); Quentin Hardy, *Using Algorithms to Determine Character*, N.Y. TIMES (July 26, 2015), https://archive.nytimes.com/bits.blogs.nytimes.com/2015/07/26/using-algorithms-to-determine-character.

99. An investigation of privacy policies and related documents found that, hypothetically, two friends who order a pizza and a movie with their devices at home might give up over 53 pieces of information about themselves. *See* Stephanie Stamm, Tripp Mickle & Jessica Kuronen, *How Pizza Night Can Cost More in Data Than Dollars*, WALL ST. J. (Apr. 10, 2018) https://www.wsj.com/graphics/how-pizza-night-can-cost-more-in-data-than-dollars.

100. *Letter to Apple and Alphabet from Commissioner Brendan Carr*, FED. COMMC'NS COMM'N (June 24, 2022), https://www.fcc.gov/sites/default/files/carr-letter-apple-and-google.pdf [hereinafter FCC Letter].

101. *See* Or Baram, *Do You Know What Personal Data Dating Apps Collect About You? (Hint: It's a Lot!)*, MINE BLOG (Feb. 13, 2022), https://blog.saymine.com/blog-1/know-what-personal-data-dating-apps-collect-about-you-february-2022; *see also* Rebecca Heilweil, *Tinder May Not Get You a Date. It Will Get Your Data.*, VOX (Feb. 14, 2020), https://www.vox.com/recode/2020/2/14/21137096/how-tinder-matches-work-algorithm-grindr-bumble-hinge-algorithms.

102. *Id.*

requested her data from Tinder received 800 pages of information—some of it quite intimate.[103]

So extensive and intimate are their data collection capabilities, some platforms have become protagonists in geopolitics and international security. Though best known for its addictive interface and amusing videos, the question of TikTok's data collection is now a matter of great power geopolitics.[104] FCC leadership recently described TikTok as an "unacceptable" risk to national security.[105] X's (formerly Twitter) combination of extensive data collection and lax cybersecurity prompted alarm.[106] Grindr, the largest LGBTQ+ social networking and dating app, was also the subject of agitation. In 2019, the Committee on Foreign Investment in the United States (CFIUS) ordered the divestment of Grindr shares owned by Beijing Kunlan Tech Co. Ltd., a Chinese gaming company.[107] Although CFIUS did not disclose any specific concerns—the inter-agency committee rarely does—speculation pointed towards blackmail risks, particularly for government contractors and personnel.[108] As large-scale mediators of information, social platforms have played roles in intelligence leaks, disinformation campaigns, and other affairs with sensitive national security implications.[109]

---

103. Judith Duportail, *I Asked Tinder for My Data. It Sent Me 800 Pages of My Deepest, Darkest Secrets*, GUARDIAN (Sept. 26, 2017), https://www.theguardian.com/technology/2017/sep/26/tinder-personal-data-dating-app-messages-hacked-sold.

104. *See, e.g.*, FCC Letter, *supra* note 100. A parallel concern is TikTok's potential to manipulate users via propaganda; *The All-Conquering Quaver*, ECONOMIST (July 9, 2022), https://www.economist.com/interactive/briefing/2022/07/09/the-all-conquering-quaver ("But there is a second, bigger fear about security, which concerns not what TikTok learns about its users, but what they learn from it.").

105. FCC Letter, *supra* note 100 ("It is clear that TikTok poses an unacceptable national security risk due to its extensive data harvesting being combined with Beijing's apparently unchecked access to that sensitive data.").

106. As noted by Charles E. Grassley, a top member of the Senate Judiciary Committee, "Take a tech platform that collects massive amounts of user data, combine it with what appears to be an incredibly weak infrastructure and infuse it with foreign state actors with an agenda, and you've got a recipe for disaster." *See* Joseph Menn, Elizabeth Dwoskin & Cat Zakrzewski, *Former Security Chief Claims Twitter Buried "Egregious Deficiencies,"* WASH. POST (Aug. 23, 2022), https://www.washingtonpost.com/technology/interactive/2022/twitter-whistleblower-sec-spam.

107. Ama Adams, Brendan Hanifin & Emerson Siegle, *Grindr, CFIUS and the National Security Risks of Dating*, LAW360 (Apr. 1, 2019), https://www.law360.com/articles/1144915/grindr-cfius-and-the-national-security-risks-of-dating.

108. *Id.* (quoting sources that concerns stemmed from blackmail risks of American officials or contractors).

109. *See, e.g.*, Tim R Samples, *My Short Life As (The Face Of) a Russian Disinformation Troll*, COLUM. JOURNALISM REV. (July 30, 2018), https://www.cjr.org/first_person/russian-troll-twitter.php ("Once pressed by the U.S. government to investigate, Twitter identified 3,814

At the heart of many digital platform business models is a vital commodity: human attention. Users who spend more time engaging with an app generate more opportunities for targeted advertising and data collection.[110] That relationship is the basis of the notion that on a "free" platform, the user is actually the product.[111] In fact, harvesting data and attention is not just a norm but the *raison d'être* of certain social apps.[112] When attention is the crux of a business model, platforms are incentivized to maximize user engagement. As an important point of reference: equity markets value social platforms, in part, as a function of their active user base and engagement.[113] Thus, a highly engaged—or, put differently, a highly addicted—user is a valuable user.

The attention-surveillance business model generates problematic incentives: more addictive platforms are more profitable platforms. Those incentives tempt platforms to deploy addictive interfaces (also called "dark patterns") to maximize user engagement.[114] Techniques—such as variable reward schedules, infinite scroll, gamification, and feedback loops—harness

---

accounts actively managed by Russian operatives and some 50,258 bots that tweeted over a million times around the election."); Jennifer Jacobs & Josh Wingrove, *US Urges Social Media to Not Share Leaked Docs in Damage Control*, BLOOMBERG (Apr. 13, 2023), https:// www.bloomberg.com/news/articles/2023-04-13/us-urges-social-media-to-not-share-leaked-docs-in-damage-control (highlighting the role of social platforms in intelligence document leaks).

110. FORBRUKERRÅDET [CONSUMER COUNCIL OF NORWAY], OUT OF CONTROL, https://web.archive.org/web/20230121151349/https://fil.forbrukerradet.no/wp-content/ uploads/2020/01/2020-01-14-out-of-control-final-version.pdf (Jan. 14, 2020) (arguing that "comprehensive tracking and profiling of consumers that is at the heart of the adtech industry are by their very nature exploitative practices"); Vikram R. Bhargava & Manuel Velasquez, *Ethics of the Attention Economy: The Problem of Social Media Addiction*, 31 BUS. ETHICS Q. 321, 322 (2021).

111. Will Oremus, *Are You Really The Product?*, SLATE (Apr. 27, 2018) https://slate.com/ technology/2018/04/are-you-really-facebooks-product-the-history-of-a-dangerous-idea.html (tracing the origins of the notion that social media users are the products in their business models).

112. *See, e.g.*, Kevin Roose, *Eight: "We Go All,"* RABBIT HOLE (June 4, 2020) https:// www.nytimes.com/2020/06/04/podcasts/rabbit-hole-qanon-youtube-tiktok-virus.html (explaining [at 31:03] that ByteDance views apps such as TikTok not as a "primary product" but instead as a vehicle for collecting data to improve artificial intelligence capabilities).

113. *See, e.g.*, Tyler Clifford, *Jim Cramer Reveals His Top Social Media Stocks*, CNBC (May 20, 2019), https://www.cnbc.com/2019/05/20/jim-cramer-reveals-his-top-social-media-stocks.html ("As long as Facebook can maintain its user and engagement numbers, this stock will remain the undisputed king of social media.") (quoting Jim Cramer, a famous equities analyst).

114. Gregory Day & Abbey Stemler, *Are Dark Patterns Anticompetitive?*, 72 ALA. L. REV. 1, 3 (2020).

the power of dopamine and neurological stimulation.[115] So-called "brain hacking" techniques are not exclusive to social platforms, nor are they universal among them, but they are prevalent enough to be considered. Combined with extraordinary scale and data sensitivies, addictive designs add another layer of differentiation between general consumer contracting and social platform TOUs.[116]

3. Bounded Choice

Digital platforms mediate almost every aspect of modern human life. Across economic, political, and social spheres, platforms organize tremendous amounts of information and human interaction.[117] The systemic importance attained by certain platforms has drawn comparisons to public utilities, "too-big-to-fail" financial institutions, essential infrastructure, and so on.[118] Network effects and the inherent scalability of software have enabled remarkable concentrations in digital markets. Data advantages can also create feedback loops (for instance, data accumulation contributes to superior user experiences) for incumbent platforms that achieve scale early.[119] As a result, the largest platforms are exceptionally large. Google, for instance, controls around 93% of the online search market.[120] As of 2021, Meta controlled three out of the five top social platforms.[121]

Individuals and organizations may find that establishing an account on a social platform is almost inevitable. Temporary outages shed light on the

---

115. See Griffin, supra note 78, at 6–14 (outlining manipulative features employed by prominent social platforms); Bjorn Lindstrom, Martin Bellander, David T. Schultner, Allen Chang, Philippe N. Tobler & David M. Amodio, *A Computational Reward Learning Account of Social Media Engagement*, 12 NATURE COMM. 1, 7 (2021) (finding that reward learning mechanisms drive human behavior on social platforms).

116. See Griffin, supra note 78, at 449 ("Understanding—and regulating—the addictive design at the core of so many Big Tech platforms is a necessary complement to work on Big Tech's antitrust, privacy, and speech issues.").

117. See, e.g., Anupam Chander, *Facebookistan*, 90 N.C. L. REV. 1807, 1809 (2012) ("Facebook increasingly records our lives, mediates our interactions, and serves as a platform for businesses, media, organizations, and even governments to engage the world.").

118. See Packin, supra note 77 and accompanying text.

119. Though dramatic in scale, platform dominance is not necessarily stable, however. See, e.g., Catherine Tucker, *Network Effects and Market Power: What Have We Learned in the Last Decade?*, ANTITRUST (2018), https://sites.bu.edu/tpri/files/2018/07/tucker-network-effects-antitrust2018.pdf (discussing cases of unstable incumbency among digital platforms).

120. *Worldwide Desktop Market Share of Leading Search Engines from January 2010 to December 2021*, STATISTA (Jan. 26, 2022), https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines (citing Google's 92.47% market share as of June 2021).

121. *Most Popular Social Networks Worldwide as of October 2021, Ranked by Number of Active Users*, STATISTA (Jan. 7, 2022), https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users.

extent of dependency on dominant platforms.[122] When the Meta platforms went down for just six hours in 2021, the consequences were serious, especially for vulnerable populations.[123] So elemental to social systems is this data, that a failure scenario at Facebook would have broad social and even cultural consequences.[124] Theories of rational behavior falter in these conditions. Because the top social platforms play such an essential role in everyday life, users are hardly facing a real choice when they click the "I agree" button.[125] Even with an activist and informed minority, opportunities to negotiate and select viable alternatives are lacking.

4. *Digital Governance*

Data is the most valuable resource in the world—dubbed the "new oil" of the digital era.[126] In the modern economy, the ownership and management of data is elemental to governance.[127] Given the scale and nature of digital platforms, TOUs play a significant role in digital governance, especially in jurisdictions that have weaker data and consumer protection laws. Platforms now play outsized roles in shaping privacy and speech rights at the global scale.[128] TOUs, in turn, are central in defining the relationship between technology and society.[129] In creating governance frameworks for the users of

---

122. Raymond Zhong & Adam Satariano, *Facebook's Apps Went Down. The World Saw How Much It Runs on Them.*, N.Y. TIMES (Oct. 8, 2021) https://www.nytimes.com/2021/10/05/technology/facebook-down-ig-down-whatsapp-down.html.

123. *See, e.g.*, Avi Asher-Schapiro & Fabio Teixeira, *Facebook Down: What the Outage Meant for The Developing World*, THOMSON REUTERS FOUND. (Oct. 5, 2021), https://news.trust.org/item/20211005204816-qzjft.

124. *See* Öhman & Aggarwal, *supra* note 77, at 5–10 (exploring consequences of failure for a variety of stakeholders).

125. AN INTRODUCTION TO ONLINE PLATFORMS AND THEIR ROLE IN THE DIGITAL TRANSFORMATION, ORG. FOR ECON. COOP. AND DEV. 13 (2019).

126. Setting aside the precision of that metaphor, it does—at the very least—reflect the vast importance of data. *The World's Most Valuable Resource Is No Longer Oil, but Data*, ECONOMIST (May 6, 2017), https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.

127. *See, e.g.*, Jennifer Daskal, *Borders and Bits*, 71 VAND. L. REV. 179, 182 (2018) ("The multinational companies that manage our data have taken on a form of international governance in ways that traditional governments can't and won't.").

128. *See* Jones & Samples, *supra* note 77, at 163–80 (outlining quasi-governmental roles of digital platforms); *see also* Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1601 (noting the "essential nature" of platforms to "modern free speech and democratic culture").

129. *See* Kim & Telman, *supra* note 4, at 754.

digital platforms, TOUs shape basic human rights such as privacy, personal security, and political participation.[130]

TOUs play a critical role in defining the relationship between technology and society. TOUs shape public discourse online by limiting some types of speech and promoting others.[131] They justify the removal of elected officials from the largest digital ecosystems on the planet.[132] Later, when a suspended or banned user—say, the former President of the United States—files a lawsuit over his removal from X (formerly Twitter), a judge looks to the platform's TOU to decide on a motion to transfer.[133] After that transfer, a California court then consults the TOU when assessing the former President's claims.[134]

Likewise, systemically important platforms wield the power to discipline other platforms, taking on quasi-regulatory functions.[135] Following the January 6 riots at the U.S. Capitol, Amazon, Apple, and Google removed Parler from their platforms for violating their terms of service.[136] Those actions are also executed under the banner of TOUs. Social issues and access to justice are shaped by TOUs as well. Whether policing impersonation claims or antisemitic content, the TOUs and policies of social platforms are determinative.[137] When a Virginia man sued Airbnb on the grounds of racial discrimination, for

---

130. *See* Dafna Dror-Shpoliansky & Yuval Shany, *It's the End of the (Offline) World as We Know it: From Human Rights to Digital Human Rights – A Proposed Typology*, 32 EUR. J. INT'L L. 1249, 1250 (2021) (highlighting concerns about online infringement of the "basic human rights of online users, such as privacy, personal security and participation on equal terms in political life").

131. *See, e.g.*, Lata Nott & Brian Peters, *Free Expression on Social Media*, FREEDOM FORUM, https://www.freedomforum.org/free-speech-on-social-media/ (last visited Mar. 10, 2024) (detailing how social media platforms moderate user content).

132. *See, e.g.*, *YouTube Says it Pulled Bolsonaro Videos for COVID-19 Misinformation*, REUTERS (July 21, 2021) (reporting on platform decisions to remove content of President Bolsonaro "for breaching their terms of use").

133. Trump v. Twitter, Inc., No. 21-22441-CIV, 2021 WL 8202673 (S.D. Fla. Oct. 26, 2021) (granting Twitter's motion to transfer an action by enforcing the TOU's forum selection clause).

134. Trump v. Twitter Inc., 602 F. Supp. 3d 1213, 1227 (N.D. Cal. 2022) (noting that the TOU "gave Twitter contractual permission to act as it saw fit with respect to any account or content for any or no reason.").

135. *See* Ian Bremmer, *The Technopolar Moment*, 100 FOREIGN AFFS. 112, 113 (2021) (observing that the most important platforms "have taken control of aspects of society, the economy, and national security that were long the exclusive preserve of the state").

136. Alex Fitzpatrick, *Why Amazon's Move to Drop Parler Is a Big Deal for the Future of the Internet*, TIME (Jan. 21, 2021), https://time.com/5929888/amazon-parler-aws.

137. *See, e.g.*, Christy Piña, *Kanye West Tweet Taken Down for Violating Twitter Rules*, BILLBOARD (Oct. 9, 2022), https://www.billboard.com/music/music-news/twitter-takes-down-kanye-west-tweet-1235153071.

instance, his claims were dismissed in court due to an arbitration agreement buried in a TOU longer than *Macbeth*.[138]

III.   DATA AND METHODOLOGY

Corpus linguistics offers a powerful set of methodologies for analyzing text and language. [139] But the emergence of corpus linguistics in legal scholarship and judicial interpretation is as controversial as it is promising.[140] The primary focus, at least recently, of legal corpus linguistics is trained on divining the ordinary meaning of certain words, often with the aim of guiding judicial interpretation.[141] We apply an entirely different set of corpus linguistics methods to an entirely different set of research questions and data. Specifically, we assess linguistic complexity, and key tendencies of social platform TOUs. This Part begins by outlining the characteristics of our dataset and our data collection process. Next, we explain our methodology. Finally, we discuss and illustrate the findings of our analysis.

A.   DATA: CHARACTERISTICS AND COLLECTION

Almost all digital platforms feature a TOU that functions as the primary user agreement. In addition, any number of supplementary terms may be incorporated into the TOU, usually via hyperlinked references.[142] Our dataset includes the primary user terms of seventy-five platforms: TOUs, privacy

---

138.   The trial court recognized the fundamental asymmetry of the situation. *See* Selden v. Airbnb, Inc., No. 16-CV-00933, 2016 WL 6476934 (D.D.C. 2016), aff'd, 4 F.4th 148 (D.C. Cir. 2021) ("While that result might seem inequitable to some, this Court is not the proper forum for policy objections to mandatory arbitration clauses in online adhesion contracts. Such objections should be taken up with the appropriate regulators or with Congress."); *see also infra* Section IV.D. (addressing the role of length in the reading difficulties of TOUs).

139.   Bernstein, *supra* note 16, at 1454 ("Corpus linguistics is a powerful methodology for analyzing the realities of language practice.").

140.   Corpus linguistics has surfaced in Supreme Court opinions, amicus briefs, and numerous scholarly works. *See supra* note 21 and accompanying text. Text analysis has also been used to detect bias in language, for instance, in letters of recommendation. *See* Charlotte S. Alexander, *Text Mining for Bias: A Recommendation Letter Experiment*, 59 AM. BUS. L.J. 1, 12–13 n.17 (2022) (describing language analysis and the corpus of recommendation letters).

141.   *See* Bernstein, *supra* note 16, at 1401 (describing the aims of legal corpus linguistics); *see also* Jennifer L. Mascott, *Who Are "Officers of the United States"?*, 70 STAN. L. REV. 443 (2018) (using corpus linguistics to determine original public meaning of the word "officer"); Lawrence M. Solan & Tammy Gales, *Corpus Linguistics as a Tool in Legal Interpretation*, 2017 B.Y.U. L. REV. 1311, 1312 (2017) (exploring "conditions in which [corpus linguistics] can be optimally employed by judges and others tasked with construing authoritative legal documents").

142.   *See, e.g., Terms of Service*, DISCORD, https://discord.com/terms (last updated Feb. 24, 2023) ("We also have a Privacy Policy, Community Guidelines, and other policies that apply to your use of our services and are incorporated into these terms.") [hyperlinks omitted].

policies, and community guidelines.[143] We focused our data collection on those three categories of user terms because they are (1) widely used across social platforms and (2) critical in defining the user's relationship to the platform.[144] Our dataset contains 195 separate texts with roughly 944,459 total words: seventy-five TOUs (504,025 total words), seventy-three privacy policies (325,793 total words), and forty-seven community guidelines (114,641 total words). In addition to those agreements and policies, we collect metadata on the platforms themselves (e.g., category, domicile) and key aspects of their TOUs (e.g., word count, governing law, dispute resolution, modification).[145]

Our data collection proceeded as follows. First, we refined the scope. We selected platforms with three key characteristics: *significant social components*, *mobile apps*, and *TOUs available in English*. For many apps, the question of significant social components is straightforward, as they are primarily (or almost exclusively) social networking platforms. However, there is no official registry of social platforms.[146] Indeed, a precise definition of "social platform" is rather difficult to pin down.[147] With the aim of building a diverse and deep dataset, we took an inclusive approach to selecting platforms.[148] As for the second criterion, we selected platforms that offer mobile apps because of their

---

143. There is more variation in the titles of policies that govern behavior on the platform. We included various forms of content, community, and acceptable behavior policies under the "community guidelines" umbrella.

144. For some apps there are even more agreements that govern the consumer-app relationship such as policies about virtual items (TikTok), cookies (Tinder), music guidelines (Snapchat), profiling (OkCupid), safety (Hoop), and so on. While these, arguably, are also relevant to defining the user-platform relationship, they are less consistently used and thus more difficult to systematically organize and assess. We count supplementary terms only when they are distinct and separate from the main TOU text. Occasionally, a privacy policy may be embedded within a TOU.

145. For a detailed description of the metadata, see *infra* Section IV.D.

146. *See* Chand Rajendra-Nicolucci & Ethan Zuckerman, *Top 100: The Most Popular Social Media Platforms and What They Can Teach Us*, KNIGHT FIRST AMEND. INST. (Jan. 22, 2021), https://knightcolumbia.org/blog/top-100-the-most-popular-social-media-platforms-and-what-they-can-teach-us ("There's no official 'registry' of website traffic that serves as a 'league table' for social media.").

147. A variety of definitions have been proposed. *See, e.g.,* danah m. boyd & Nicole B. Ellison, *Social Network Sites: Definition, History, and Scholarship*, 13 J. COMPUT.-MEDIATED COMM. 210, 211 (2008) (articulating an early definition of "social network"). Making matters more convoluted, many websites and applications have social features that are secondary or supplementary to the primary service, such as Venmo.

148. We included dating applications and some prominent fintech applications with social features, for instance. *See infra* notes 153–155 (describing the selection inputs). Whether or not dating platforms "count" as social platforms is the subject of debate. *Compare* Rajendra-Nicolucci & Zuckerman, *supra* note 146 (discussing the question of whether or not dating apps are social platforms) *with supra* Section II.B (comparing and contrasting business models of dating apps with other social platforms).

enhanced data collection capacities.[149] Finally, we limited our selection to platforms with TOUs available in English because some of our methods are designed specifically for the English language.[150]

Second, we built the dataset. We began by adding market-leading platforms—those with the most downloads, largest reported user bases, and highest in popularity.[151] Selecting platforms with the most downloads or largest active user bases is usually straightforward, but data is limited outside of the top apps. Data on active users is contested even on the most visible of publicly traded platforms.[152] In addition to rankings by downloads and active users, we considered the social media mapping project by the Knight First Amendment Institute at Columbia University, which indexes "popularity" among social platforms.[153] We included the top thirty platforms from the Knight popularity list. Using multiple inputs to build a list of prominent social apps generated a more comprehensive, diverse dataset.

We further diversified the dataset with additional categories: the top dating apps, prominent "alt-tech" social networking platforms,[154] and two fintech platforms with meaningful social components.[155] Some definitions of social media exclude dating apps.[156] However, consistent with Michael Rustad and

---

149. Mobile applications have greater data gathering capabilities—and, thus, greater implications for consumer rights and privacy—than purely web-based platforms. *See supra* notes 89–95. Thus, we excluded 4chan, which does not offer a proprietary mobile application.

150. This criterion, unfortunately, precluded some interesting platforms from the dataset such as Douyin (similar to TikTok but available in China) and Taringa! (a social platform based in Argentina).

151. *See, e.g., Most Popular Social Networks Worldwide as of January 2022, Ranked by Number of Monthly Active Users*, STATISTA (June 21, 2022), https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users (ranking seventeen social platforms by monthly active users); *Leading Social Media Apps Worldwide in 2021, By Downloads*, STATISTA (Jan. 27, 2022), https://www.statista.com/statistics/1284900/top-social-media-apps-worldwide-by-downloads (ranking ten social platforms by downloads).

152. *See* Sarah E. Needleman, *Behind Fake-Account Issue That Elon Musk Cited in Calling Twitter Deal 'On Hold'*, WALL ST. J. (May 15, 2022), https://www.wsj.com/articles/behind-fake-account-issue-that-elon-musk-cited-in-pausing-twitter-deal-11652612403 (pointing out that "[s]pam and fake accounts are an industrywide challenge.").

153. *See* Rajendra-Nicolucci & Zuckerman, *supra* note 146.

154. Although alt-tech platforms have relatively smaller user bases than incumbent social networking platforms, we favor including them to diversify the dataset. *See* Ethan Zuckerman & Chand Rajendra-Nicolucci, *Deplatforming Our Way to the Alt-Tech Ecosystem*, KNIGHT FIRST AMEND. INST. (Jan. 11, 2021), https://knightcolumbia.org/blog/deplatforming-our-way-to-the-alt-tech-ecosystem (explaining the alt-tech ecosystem).

155. In the fintech category, we selected Venmo and Public but not PayPal or Robinhood because the former have more prominent social functions in their interfaces.

156. *See* Rajendra-Nicolucci & Zuckerman, *supra* note 146 ("Dating sites were more difficult, but in the end, we decided that they were more akin to platforms like Uber which

Thomas Koenig, we favor including the dating category for this study.[157] Dating apps manifest key functions and characteristics of social platforms that make them especially sensitive for consumers and public interests.[158] Thus, we included several of the most popular dating apps in the United States and worldwide.[159] Once we selected the seventy-five platforms, we manually scraped the text of their TOUs and supplementary terms.[160]

We also collected metadata on key terms and characteristics of the platforms and their terms.[161] Our metadata analysis includes several variables: app domicile, word counts, user base, governing law, dispute resolution (arbitration or litigation, arbitration opt-outs, and class waivers), and app category. *App domicile* indicates the home country or headquarters of the platform. *Word count* refers, simply, to the number of words in a TOU or the supplementary terms. *User base* is the number of active users on a platform. *Governing law* is the applicable law, as specified in the TOU. *Dispute resolution* variables include *arbitration* or *litigation*, plus additional binary (yes/no) inputs for TOUs with arbitration: *opt-outs* and *class waivers*. *App category* refers to the market category of the platform.[162]

B.   METHODOLOGY: METRICS AND COMPUTATION

Readability is frequently defined as the ease with which a text can be read and understood in terms of its linguistic features.[163] Text length, syntactic structures, lexical features, text cohesion, paragraph size, sentence length,

---

operate two-sided marketplaces—i.e., Tinder is a matchmaking platform with social features.").

157.   Like this Article, Rustad and Koenig study a dataset of social platform TOUs. *See* Rustad & Koenig, *supra* note 4, at 1442–43.

158.   *See supra* notes 101–103, 107–108 and accompanying text.

159.   *See Most Popular Dating Apps in the United States in July 2023, By Number of Monthly Downloads*, STATISTA (Aug. 8, 2023), https://www.statista.com/statistics/1238390/most-popular-dating-apps-us-by-number-of-downloads.

160.   To prepare the dataset for analysis with R, a series of computational algorithms and manual formatting was employed for data preparation and formatting utilizing XML tagging, the Unix command line, and R packages dplyr and the tidyverse. *See generally* HADLEY WICKHAM & GARRETT GROLEMUND, R FOR DATA SCIENCE: IMPORT, TIDY, TRANSFORM, VISUALIZE, AND MODEL DATA (1st ed. 2017).

161.   For the results of our metadata analysis, see *infra* Section IV.D.

162.   We use an adapted version of the approach to categorizing social platforms developed by the Knight First Amendment Institute. *See* Rajendra-Nicolucci & Zuckerman, *supra* note 146.

163.   *See, e.g.*, Scott Crossley, Stephen Skalicky & Mihai Dascalu, *Moving Beyond Classic Readability Formulas: New Methods and New Models*, 42 J. RSCH. IN READING 541, 543 (2019) ("Text readability is best defined as the ease with which a text can be read and understood in terms of the linguistic features found within a text.").

syllable structures, and word type all factor into linguistic difficulty.[164] Recognizing the multivariate reality of reading difficulty, our analysis employs diverse points of measurement. We supplement traditional readability tests with measures of linguistic complexity that consider lexical and syntactic structures. All together, we apply five metrics to the dataset: two traditional readability formulas, an index that measures the syntactic complexity of verb structures, a composite score of syntactic complexity that weighs nineteen separate nominal structures, and a lexical diversity test.

1.   *Traditional Metrics*

Our calculations include two traditional readability metrics, the Flesch Reading Ease (FRE) test[165] and the Flesch-Kincaid (F-K) test.[166] FRE was developed by Rudolph Flesch in the 1940s.[167] Decades later, the F-K test, designed by Flesch and John P. Kincaid, was tested on U.S. Navy technical personnel.[168] FRE results range on a scale of zero to one hundred. The higher the FRE score, the more readable the text is supposed to be. Although the inputs are the same, the coefficients of the F-K[169] test differ from the FRE[170] formula, producing results on a scale that indicates the grade level(s) required for reading ease and understanding. Thus, the lower the F-K score, the more readable the text is supposed to be. F-K is a reformulation of FRE—not a fundamentally different test.[171]

Generally, readability formulas are based on the length of sentences and words within a text.[172] *Words per sentence* functions as a proxy for syntactic

---

164.   *See id.* at 542–43 (noting a variety of factors that contribute to linguistic difficulty); Edward G. Fichtner, *Measuring Syntactic Complexity: The Quantification of One Factor in Linguistic Difficulty*, 13 DIE UNTERRICHTSPRAXIS (TEACHING GERMAN) 67, 67–70 (1980) (same).

165.   *See* Rudolph Flesch, *A New Readability Yardstick*, 32 J. APPLIED PSYCH. 221 (1948).

166.   *See* J. PETER KINCAID, ROBERT P. FISHBURNE JR., RICHARD L. ROGERS & BRAD S. CHISSOM, DERIVATION OF NEW READABILITY FORMULAS (AUTOMATED READABILITY INDEX, FOG COUNT AND FLESCH READING EASE FORMULA) FOR NAVY ENLISTED PERSONNEL, INST. FOR SIMULATION & TRAINING (1975).

167.   *See* Rustad & Koenig, *supra* note 4, at 1459 n.150 (referencing the origins of the FRE test).

168.   *See* Kincaid et al., *supra* note 166.

169.   The F-K formula is computed as $(0.39 \times ASL) + (11.8 \times ASW) - 15.59$. ASL represents average sentence length while ASW represents average number of syllables per word.

170.   The FRE formula is computed as $06.835 - (1.015 \times ASL) - (84.6 \times ASW)$.

171.   Ian Gallacher, *"When Numbers Get Serious": A Study of Plain English Usage in Briefs Filed Before the New York Court of Appeals*, 46 SUFFOLK U. L. REV. 451, 460 (2013) ("The [F-K] test is a reformulation of the [FRE] Score test that expresses its result in terms of the grade level a hypothetical reader should have achieved before the selected passage would be readable.").

172.   Rustad & Koenig, *supra* note 4, at 1458–61 (same); Crossley et al., *supra* note 163, at 542 ("Generally, these formulas rely on superficial text-based features to assess readability

complexity; *syllables per word* acts as a proxy for lexical difficulty.[173] Limitations aside, traditional formulas have some advantages: they are readily available, simple to use, and easily scalable. Even common Microsoft products like Word and Outlook can execute FRE and F-K tests. Such advantages might help explain why FRE and F-K remain widely used, including by educational institutions and military agencies.[174] A further advantage is specific to this Article: using traditional formulas in our analysis allows us to compare our results with previous TOU studies.[175]

That said, traditional readability formulas face major criticisms. Despite their billing as "readability" tests, traditional formulas—such as FRE, F-K, and several others—are limited by their narrow scope of inputs.[176] At best, they are simplistic and outdated. At worst, they lack construct validity[177] and perform poorly at their purported function of predicting readability.[178] Linguists have largely abandoned the traditional readability formulas, relying instead on more advanced metrics for evaluating the difficulty of a text. For the purposes of our study, we view traditional formulas as helpful points of reference— particularly in concert with more robust linguistic metrics—but inadequate as a standalone methodology. Accordingly, our analysis also includes syntactic complexity and lexical diversity.

---

including the number of words per sentence, which is meant to act as a proxy for syntactic complexity, and the number of characters per word, which is meant to act as a proxy for lexical difficulty.").

173.   Crossley et al., *supra* note 163, at 542.

174.   *See* Kathy Conklin, Richard Hyde & Fabio Parente, *Assessing Pain and Intelligible Language in the Consumer Rights Act: A Role for Reading Scores?*, 39 LEGL. STUDIES 378, 385–87 (2019) (outlining examples); *see also* S.C. Code Ann. § 37-4-105(B) (2015) (requiring certain consumer-oriented insurance disclosures to score "no higher than seventh grade on the Flesch-Kincaid readability test").

175.   *See infra* Table 1 (comparing FRE and F-K scores from our corpus with previous studies).

176.   That limitation also applies to other classic readability formulas such as FOG and SMOG. *See* Conklin et al., *supra* note 174, at 385–87 (reviewing prominent readability formulas).

177.   Scott A. Crossley, Stephen Skalicky, Mihai Dascalu, Danielle S. McNamara & Kristopher Kyle, *Predicting Text Comprehension, Processing, and Familiarity in Adult Readers: New Approaches to Readability Formulas*, 54 DISCOURSE PROCESSES 340, 342 (2017) ("[Traditional readability] formulas are generally not based on theories of reading or comprehension but rather rely on statistical correlations to develop predictive power.").

178.   *See* Crossley et al., *supra* note 163, at 557 (finding that more advanced methods outperformed traditional readability formulas); *see also* Scott P. Ardoin, Shannon M. Suldo, Joseph Witt, Seth Aldrich & Erin McDonald, *Accuracy of Readability Estimates' Predictions of CBM Performance*, 20 SCHOOL PSYCH. Q. 1, 15 (2005) (finding low accuracy among readability formulas).

2. *Syntactic Complexity*

Syntax is another distinctive internal linguistic factor at play in text readability.[179] Syntax refers to the ways that words may be combined to create meaningful units of language. The grammatical categories and linguistic patterns involved in a phrase or sentence, including verbs and nominals, form syntactic structures. Thus, *syntactic complexity* reflects the range and complexity of language forms in a given text.[180] Compound sentences, embedded structures, and modifying clauses, for instance, make sentences more syntactically complex. The simple sentence ("Tracking helps us.") becomes more syntactically complex by adding further grammatical elements. For example, the sentence ("Among other things, tracking helps us.") is more syntactically complex with a prepositional modifying phrase ("among other things").

Linguists often measure syntactic complexity with specific or composite scores that quantify structures and categories of syntax.[181] Linguistic studies have considered syntactic complexity in diverse settings, from legal texts to spoken language.[182] To assess syntactic complexity in our dataset, we rely on two analytical tools: Fichtner's C index[183] and the Tool for the Automated Analysis of Syntactic Sophistication and Complexity (TAASSC).[184] Whereas the C index measures the complexity of verb structures, we use TAASSC to quantify the complexity of nominal structures.[185] Thus, the results we generate with Fichtner's C and TAASSC are complementary. Capturing the syntactic

---

179. *See* John Brennan, Yuval Nir, Uri Hasson, Rafael Malach, David J. Heeger & Liina Pylkkanen, *Syntactic Structure Building in the Anterior Temporal Lobe During Natural Story Listening*, 120 BRAIN & LANGUAGE 163 (2012); *see also* Miloš Stanojević, Shohini Bhattasali, Donald Dunagan, Luca Campanelli, Mark Steedman, Jonathan R. Brennan & John Hale, *Modeling Incremental Language Comprehension in the Brain with Combinatory Categorical Grammar*, PROC. OF THE WORKSHOP ON COGNITIVE MODELING & COMPUTATIONAL LINGUISTICS 23 (2021).

180. *See* Larsson & Kaatari, *supra* note 23, at 1.

181. *See*, *e.g.*, Eugène Mollet, Alison Wray, Tess Fitzpatrick, Naomi R. Wray & Margaret J. Wright, *Choosing the Best Tools for Comparative Analyses of Texts*, 15 INT'L J. CORPUS LINGUISTICS 429, 443–47 (2010) (discussing linguistic metrics that measure grammatical sophistication).

182. *See*, *e.g.*, Tatian Tkacukova, *Forensic Linguistics and Language and the Law*, *in* AN INTRODUCTION TO APPLIED LINGUISTICS (2019); *see also* Zamanian & Heydari, *supra* note 23.

183. *See generally* Fichtner, *supra* note 164.

184. *See* Kristopher Kyle, *Measuring Syntactic Development in L2 Writing: Fine Grained Indices of Syntactic Complexity and Usage-Based Indices of Syntactic Sophistication* (May 9, 2016) (Ph.D. dissertation, Georgia State University), https://doi.org/10.57709/8501051.

185. Nominal groups, generally, are grammatical units that can be used as nouns. The noun phrase ("our users") includes the possessive dependent ("our") with the main noun ("users"). This noun phrase can be made more complex by adding an adjectival modifier ("diverse") as in ("our diverse users").

complexity of both verb and nominal structures enhances the diversity and scope of our results.

a)   Fichtner's C Index: Verb Structures.

Verbs are an important indicator of syntactic complexity. Fichtner's C index approaches syntactic complexity by measuring the density of verb structures: the number of verbs per sentence scaled by average sentence length.[186] Put another way, Fichtner's C is calculated as "the number of word tokens times the number of lexical verb tokens, divided by the square of the number of sentences."[187] Thus, the formula is operationalized across a sample text as a proportion of verbs per sentence relative to the number of words per sentence. That operation allows for accurate comparisons across texts of varying lengths.

A key insight of Fichtner's theory: the syntactic complexity of a text is driven by the density of lexical verbs within sentences. In other words, Fichtner saw syntactic complexity as a function of the number of lexical verbs per sentence in relation to the length of those sentences.[188] That insight is simple but effective. In a comprehensive study of 381 metrics derived from approximately 200 analytical tools for comparative linguistic analysis, Fichtner's C index was found highly effective.[189] In fact, the Fichtner's C index was identified as the "most promising" of all the tools studied for evaluating linguistic complexity.[190]

Although Fichtner's C requires a minimum word count for reliable score output, the following short sentences help to illustrate differences in the syntactic complexity of verb structures.[191] The simple sentence ("The cat sleeps") would contribute to a lower overall Fichtner's C score. By comparison, the sentence ("The cat sleeps and eats and chases birds all day") would register a higher overall score.

Complex, technical sentences with elaborate verb structures score highest in Fichtner's C index. Our results indicate that such sentences are relatively

---

186.   *See* Fichtner, *supra* note 164, at 71.

187.   Mollet et al., *supra* note 181, at 445.

188.   *Id.*

189.   *Id.*

190.   *Id.* (characterizing Fichtner's C as "most promising, on the basis of its relative simplicity, its mathematical robustness, and its correlation with other measures").

191.   A sample should contain at least 500 words for measurement with this index. *See* Fichtner, *supra* note 164, at 71.

common in TOUs.[192] For instance, the indemnification clause from Twitch's TOU registers an exceptionally high complexity score:[193]

> To the fullest extent permitted by applicable law, you agree to indemnify, defend, and hold harmless Twitch, its affiliated companies, and each of our respective contractors, employees, officers, directors, agents, third-party suppliers, licensors, and partners (individually and collectively, the "Twitch Parties") from any claims, losses, damages, demands, expenses, costs, and liabilities, including legal fees and expenses, arising out of or related to your access, use, or misuse of the Twitch Services, any User Content you post, store, or otherwise transmit in or through the Twitch Services, your violation of the rights of any third party, any violation by you of these Terms of Service, or any breach of the representations, warranties, and covenants made by you herein.

This indemnification clause is a characteristic example of the grammatical complexity and linguistic patterns within TOUs. It combines length and lexical difficulty with elevated syntactic complexity.

b) TAASSC: Nominal Structures.

We use TAASSC to measure the complexity of nominal structures. TAASSC works by counting and tagging different syntactic structures and their averages across texts of interest.[194] Crucially, rather than scores based on the number of words, TAASSC uses grammatical relations to calculate syntactic complexity.[195] Namely, TAASSC counts the number of dependents per governing phrase type. Take, for instance, a sentence ("You retain your rights to your content.") that includes two dependent nominal phrases of the governing verb *retain*. The dependents of the verb are the nominal subject, ("You") and the direct object ("rights"). These nominal phrases include the following dependents: the possessive adjective "your" (which occurs twice) and one prepositional phrase. The average number of adjectival dependents per nominal is 1.0 (two divided by two). By making computations based on grammatical relations, TAASSC avoids over-indexing for structures that have higher average word counts (for instance, prepositional phrases versus adjectival modifiers).

---

192. *See infra* Section IV.B (comparing Fichtner's C results between our dataset and other genres of English).

193. *Terms of Service*, TWITCH, https://www.twitch.tv/p/en/legal/terms-of-service (last updated Oct. 27, 2023).

194. TAASSC uses python to parse texts and collect the averages for over 30 different types of clause and phrase structures. TAASSC outputs variables of syntactic, clause, and phrase complexity. We use TAASSC version 1.3.8 and python version 2.7.

195. *See* Kyle, *supra* note 184, at 54–55.

Embedding is another key phenomenon associated with text complexity.[196] Embedding refers to the insertion of grammatical units into additional units.[197] Examples of insertion include the placement of phrases, dependents, or other clause types within sentences, clauses, or phrases. Embedding and complex noun phrase structures are key characteristics of academic writing, for instance.[198] Previous research has also found greater use of these patterns in the writing of more fluent English language learners—an indication of higher writing sophistication.[199] Embedding in syntactic structures is also associated with greater difficulty in terms of cognitive processing and reading ease.[200]

TAASSC examines four main categories of syntactic sophistication and complexity, with over thirty indices of clausal and phrasal complexity.[201] We focus our analysis on the results of noun phrase (NP) elaboration, the composite score of all nineteen TAASSC noun phrase types and embedding indices. Specifically, NP elaboration measures grammatically embedded elements, including the number of dependents per noun phrase type, determiners, adjectives, prepositions, and verbal modifiers of nominals.[202] Averages for each type and dependents per type are calculated with TAASSC and then combined for the NP elaboration results.[203] Annex 1 provides further illustration of inputs in the NP elaboration score.

---

196. *See* María Belén Díez-Bedmar & Pascual Pérez-Paredes, *Noun Phrase Complexity in Young Spanish EFL Learners' Writing: Complementing Syntactic Complexity Indices with Corpus-Driven Analyses*, 25 INT'L J. OF CORPUS LINGUISTICS 4, 8 (2020).

197. *See, e.g.*, LISE FONTAINE, ANALYSING ENGLISH GRAMMAR: A SYSTEMIC-FUNCTIONAL INTRODUCTION 23 (2012) (explaining how embedding can increase complexity and providing examples).

198. *See* Douglas Biber & Bethany Gray, *Grammatical Change in the Noun Phrase: The Influence of Written Language Use*, ENGLISH LANGUAGE & LINGUISTICS 223, 223 (2011) (noting that academic writing styles rely "heavily on nominal structures, with extensive phrasal modification and a relative absence of verbs"); *see also* Kyle, *supra* note 184, at 16.

199. *See* Kyle, *supra* note 184, at 34; Larsson & Kaatari, *supra* note 23, at 5.

200. *See* Martínez et al., *supra* note 23, at 2, 6; Arthur C. Graesser, Danielle S. McNamara, Zhiqang Cai, Mark Conley, Haiying Li & James Pennebaker, *Coh-Metrix Measures Text Characteristics at Multiple Levels of Language and Discourse*, 115 ELEMENTARY SCHOOL J. 210, 213 (2014); Haeran Jae, *Cognitive Load and Syntactic Complexity of Printed Advertisements: Effects on Consumers' Attitudes*, 21 MARKETING MGMT. J. 152, 153, 157–08 (2011).

201. *See* Kyle, *supra* note 184, at 51–56.

202. *See* Díez-Bedmar & Pérez-Paredes, *supra* note 196, at 9; Kyle *supra* note 184, at 71–72.

203. Noun phrase types include passive nominal subjects like ("your account") in the sentence, "Your account was terminated." Another passive example is the nominal complement ("her notice") in the sentence, "The individual was given her notice." In addition to averages of noun phrase types, NP elaboration measures the number of dependents per type. A higher number of dependents yields increased NP complexity.

Complex nominal structures contribute to higher NP elaboration scores. Previous research using TAASSC has focused predominately on language learners and educational contexts.[204] Those studies indicate that English learners incorporate greater NP complexity in their writing as they increase their language proficiency.[205] In other words, NP complexity and writing sophistication are closely associated. Our results indicate that complex syntactic structures are especially prevalent in TOUs.[206] For instance, this sentence from a "user-generated content" clause in TikTok's TOU would contribute to a higher NP elaboration score:[207]

> If you only own the rights in and to a sound recording, but not to the underlying musical works embodied in such sound recordings, then you must not post such sound recordings to the Services unless you have all permissions, clearances from, or are authorized by, the owner of any part of the content to submit it to the Services.

3.   Lexical Diversity

Lexical diversity metrics are another way to consider the complexity of a text. Previous studies have considered lexical richness in the legal context, including advocacy in favor of plain English and less technical jargon in legal texts.[208] Metrics for evaluating lexical diversity are computed based on the type-token ratio: the total number of different words (i.e., types) divided by the total number of words (i.e., tokens) in the dataset.[209] More traditional measures of lexical diversity rely on a simple calculation of vocabulary size divided by total number of words. The problem with the traditional type-token approach is that the results are affected by the length of a text.[210] Shorter texts, for instance, may have artificially high type-token ratios because the denominator is small.

---

204.   *See generally* Díez-Bedmar & Pérez-Paredes, *supra* note 196; Kristopher Kyle & Scott A. Crossley, *Measuring Syntactic Complexity in L2 Writing Using Fine-Grained Clausal and Phrasal Indices*, 102 MODERN LANG J. 333 (2018).

205.   *See, e.g.*, Díez-Bedmar & Pérez-Paredes, *supra* note 196, at 5, 26.

206.   *See infra* Section IV.B (comparing TAASSC results between our dataset and other genres of English).

207.   *Terms of Service*, TIKTOK, https://www.tiktok.com/legal/terms-of-service-us (last updated Nov. 2023).

208.   *See generally* Anna Sobota, *The Plain Language Movement and Modern Legal Drafting*, 20 COMP. LEGILINGUISTICS 50 (2014); TERESA FANEGO & PAULA RODRÍGUEZ-PUENTE, CORPUS BASED RESEARCH ON VARIATION IN ENGLISH LEGAL DISCOURSE (2019).

209.   *See* VACLAV BREZINA, STATISTICS IN CORPUS LINGUISTICS 57 (2018).

210.   *See* Covington & McFall, *supra* note 23, at 94 ("The problem is that the TTR of a text sample is affected by its length; obviously, the longer the text goes on, the more likely it is that the next word will be one that has already occurred.").

We use the Moving Average Type-Token Ratio (MATTR) score to measure lexical diversity.[211] Linguists have proposed variations on the type-token ratio, such as MATTR, to accommodate for the effects of dataset size and text length. MATTR mitigates the text length problem by calculating the ratios on a moving average window across the full length of the text sample, which normalizes the results.[212] As a result, MATTR measures lexical diversity but avoids the effects of text length and statistical assumptions.[213] We utilize the R package quanteda to compute the MATTR results across different subsections of the dataset. MATTR is a useful way to evaluate lexical aspects of text complexity because it calculates vocabulary richness across all possible subsets of the data.

4. *Comparative Analysis*

Another point of differentiation in our methodology is that we perform comparative analysis across diverse genres of English. Our comparative analysis underscores the characteristics of app-based consumer contracts by establishing external reference points. Specifically, we compare the results from our dataset with other corpora that represent genres of English: a broad and multi-genre collection of modern American English (the Brown Corpus) and a canon of iconic English literature (the Jane Austen Corpus). These comparisons add context and texture to our results.

Nelson Francis and Henry Kučera compiled the Brown University Standard Corpus of Present-Day American English, commonly referred to as the Brown Corpus.[214] The Brown Corpus is the first computerized collection of American English and remains a widely used dataset. It consists of just over 1 million words of carefully sampled texts from fifteen different genres of American English from the 1960s.[215] The corpus contains a wide range of style and prose. Fiction and news media were included in the corpus, but forms like verse and drama were excluded because they present problems for consistent linguistic analysis.[216] The stated aim for selection was representativeness—as

---

211. *Id.*

212. MATTR utilizes an overlapping window throughout the data to compute the type-token ratio across each window; the results are then averaged to for each window calculation. *Id.* at 96.

213. *Id.* ("The mean of all these TTRs is a measure of the lexical diversity of the entire text and is not affected by text length nor by any statistical assumptions.").

214. W. Nelson Francis & Henry Kučera, *Brown Corpus Manual*, http://icame.uib.no/brown/bcm.html (1979).

215. The Brown Corpus consists of 500 texts, each consisting of just over 2,000 words. *Id.*

216. The texts were sampled from 15 different text categories: press reports (44 texts), press editorials (27 texts), press reviews (17 texts), religious (17 texts), skill and hobbies (36

opposed to a subjective quality of excellence.[217] Comparative study was an intended use of the corpus.[218]

As another point of comparison, the Jane Austen Corpus is a notable source of literature texts, at just over 854,000 words. It is composed of six novels written by Jane Austen: *Mansfield Park*, *Sense and Sensibility*, *Emma*, *Pride and Prejudice*, *Northanger Abbey*, and *Persuasion*. Jane Austen is a world-renowned author, famous for her distinctive writing style and witty portrayal of social norms.[219] Austen's novels were compiled by Project Gutenberg, a free online library founded in 1971 with the mission of preserving literary and other genres of writing.[220] We obtained those novels from Project Gutenberg, then processed the raw text following Julia Silge's methodology for the janeaustenr package and dataset.[221]

5. *Calculations*

Each of the analytics is calculated across different divisions of data. Our analysis includes word, sentence, clause, and document-level metrics. At the word and sentence divisions, our metrics consider the linguistic complexity of the dataset. At the document level, we examine the TOUs and supplementary terms in our dataset, drawing comparisons between TOUs and the various categories of agreements (e.g., privacy policies versus TOUs versus community guidelines).[222] We also examine linguistic complexity at a clause-specific division, comparing arbitration clauses with other provisions. Finally, for the sake of context and comparative analysis, we apply and compare the dataset with other external corpus datasets.[223]

As for computational methods, we implement algorithms and functions using the R Programming Language for Statistics and Graphics with R

---

texts), popular lore (48 texts), belles-lettres (75 texts), government and law (30 texts), academic/learned (80 texts), general fiction (29 texts), mystery fiction (24 texts), science fiction (6 texts), adventure fiction (29 texts), romance fiction (29) texts, and humor (9 texts). *Id.*

217. *Id.*

218. *Id.* (expressing hope that the corpus will be used in comparative studies).

219. *See generally* MARILYN BUTLER, JANE AUSTEN, OXFORD UNIVERSITY PRESS (2010)*; see also* CLAIRE TOMALIN, JANE AUSTEN: A LIFE (2007).

220. *Background, History and Philosophy of Project Gutenberg*, PROJECT GUTENBERG, https://www.gutenberg.org/about/background/index.html (last visited Aug. 2, 2022).

221. *See* Julia Silge, janeaustenr: Jane Austen's Complete Novels, v. 0.1.5 (2017).

222. *See infra* Sections IV.A–D.

223. *See supra* notes 214–220 and accompanying text.

packages quanteda,[224] tidytext,[225] and polmineR.[226] Our data processing also involves submission of the data to part-of-speech tagging, tokenization, and lemmatization through the spacyr and udpipe packages and corresponding functions.[227] Developed at Bell Labs, R is a language and environment for statistical computing and graphics.[228] R offers tools at every stage of data processing: cleaning, organizing, formatting, analyzing, and visualizing. R is also an open-source language, so it is free to use and has a vibrant worldwide community of users.[229] R compiles and runs on a wide variety of platforms including Unix, Windows, and macOS. Finally, we visualize the data with Tableau Public.

IV.   FINDINGS AND DISCUSSION

In this Part, we illustrate and discuss the results. We begin with findings from traditional readability formulas. We then illustrate our findings for syntactic complexity: the results of our Fichtner's C index and TAASSC scores. Following that, we explain our findings on lexical richness, the MATTR scores. Finally, we outline and discuss the results of our metadata analysis. A table of selected results and metadata across all the individual platforms in our dataset is included in Annex 2.[230]

A.   TRADITIONAL READABILITY METRICS

This Section IV.A explains and illustrates the results of our calculations with traditional readability metrics, including some comparisons with previous studies.

---

224.   *See* Kenneth Benoit, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, & Akitaka Matsuo, *quanteda: An R Package for the Quantitative Analysis of Textual Data*, 3 J. OPEN SOURCE SOFTWARE 744 (2018).

225.   *See* Julia Silge & David Robinson, *tidytext: Text Mining and Analysis Using Tidy Data Principles in R*, 1 J. OPEN SOURCE SOFTWARE 37 (2016).

226.   *See* Andreas Blætte & Christoph Leonhardt, polmineR: Verbs and Nouns for Corpus Analysis, v. 0.8.0 (2019).

227.   *See* Kenneth Benoit & Akitaka Matsuo, spacyr: Wrapper to the 'spaCy' 'NLP' Library, v. 1.2.1 (2022); *see also* Jan Wijffels, Milan Straka, & Jana Straková, udpipe: Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the 'UDPipe' 'NLP' Toolkit, v. 0.8.9 (2022).

228.   *What is R?*, R, https://www.r-project.org/about.html (last visited Oct. 28, 2023).

229.   R is a different implementation of the S language and is maintained internationally by a team of developers through the Comprehensive R Archive Network (CRAN). *The R Project for Statistical Computing*, R-PROJECT, https://www.r-project.org (last visited Aug. 2, 2022).

230.   Annex 2 illustrates, for each platform: app category, FRE, MATTR, Fichtner's C, TAASSC, word count, domicile, and governing law.

1. *Flesch Reading Ease (FRE)*

FRE scores represent the difficulty of reading and understanding texts. As noted in our methodology explanation, FRE scores are calculated based on average sentence length and average syllables per word.[231] Figure 1 (below) displays the results of our FRE calculations. The higher the FRE score, the more readable the text is. FRE scores above sixty are considered to meet a "plain English" standard.[232] By way of reference, *Reader's Digest* scores around sixty-five whereas *Time* magazine scores about fifty-two.[233] In our calculations, the Austen Corpus scores over sixty and the TOUs in our dataset average just over thirty.

Figure 1: FRE Scores

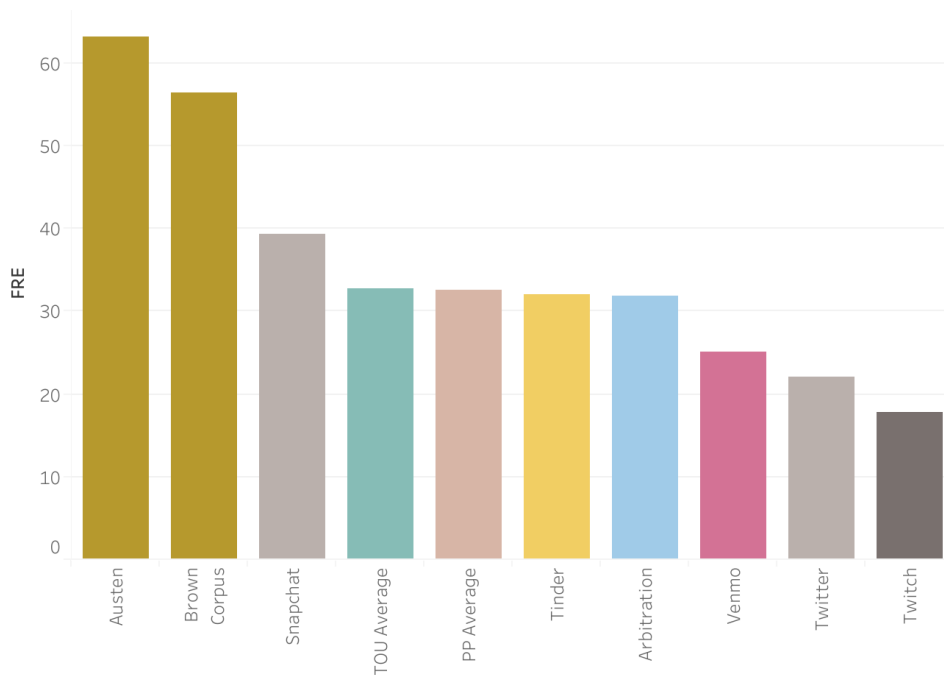Figure 1 illustrates the average FRE scores for the TOUs and privacy policies. This figure also illustrates scores for the Austen Corpus, the Brown Corpus, and some individual platform TOUs. These results suggest that most

---

231. *See supra* notes 172–173.

232. *See* Karen A. Schriver, *Plain Language in the U.S. Gains Momentum: 1940-2015*, 60 IEEE TRANSACTIONS ON PROFESSIONAL COMMC'N 343, 351 (2017).

233. *Id.*

TOUs are incomprehensible to a broad audience.[234] FRE results for all seventy-five TOUs produced an average of 32.69 points and a median score of 32.63. No TOU in the dataset scored above forty-seven points. The Swarm TOU, for instance, registered the highest individual FRE score but still falls well short of plain language standards. And the least readable TOUs are very unreadable: thirty-one platforms are in the lowest range of FRE scores (zero to thirty), which would require the completion of undergraduate and potentially some graduate level education.[235] Tantan, for instance, the least readable TOU in the dataset, has an FRE score of 15.5. Arbitration clauses register especially low scores, slightly lower than the TOU average. The FRE results indicate major differences between legal texts and other genres of language—a common thread throughout our results.[236] Curiously, however, the FRE formula does not detect a major difference between privacy policies and TOUs.[237]

2. Flesch-Kincaid (F-K) Test

F-K score results range on a scale of zero to eighteen, which approximates the years of education required to understand the text. Thus, the lower the score, the more readable the text is. Because the F-K test operates on the same inputs and illustrates the same characteristics as the FRE test, we do not discuss the F-K results separately at length. Our abbreviated F-K findings: Across TOUs, the median and mean results of the F-K calculations were 15.83 and 15.76, respectively, which indicate that at least some undergraduate coursework is required to understand the average TOU in our dataset. As external points of reference, our F-K results indicate that understanding the Jane Austen Corpus requires ninth grade education and the Brown Corpus requires upper-level high school education.

As indicated by Table 1 (below), the readability scores of online TOUs appear to have declined sharply. In the case of FRE scores, that means lower numeric values. The drop is consistent across both of the more recent datasets. For the F-K results, the decline in readability registers as a higher score, which suggests that the years of education needed to understand TOUs increased by

---

234. According to Zamanian and Heydari, the estimated percentage of U.S. adults at the 7th grade reading level is about 88%, but the rates drop quickly at higher levels of reading ability. For instance, the percentage of adults at eighth to ninth grade reading levels is 83% and at college level is 33%. *See* Zamanian & Heydari, *supra* note 23, at 45.

235. *Id.* at 44–45.

236. *See infra* Figures 2–4.

237. The median FRE score for TOUs is 32.6 while the median score for privacy policies is 32.2. We discuss this divergence at length in Section IV.B. However, the syntactic complexity scores diverge significantly. *See infra* notes 245–249.

about four years. TOUs were already quite unreadable when they were measured by Rustad and Koenig around 2014.[238] Unlike Rustad and Koenig, which focused on social platforms, the Benoliel and Becher dataset in 2019 was a general TOU dataset. Our comparison across datasets suggests TOUs have grown more complex in recent years, but more research is needed to fully explore this trend.

Table 1: Median FRE and F-K Scores for TOUs

Study	FRE	F-K
Rustad & Koenig (2014)	49	11
Benoliel & Becher (2019)	34.2	14.9
Samples, Ireland, Kraczon (2024) (this Article)	32.6	15.83

B.   SYNTACTIC COMPLEXITY METRICS

This Section IV.B explains and illustrates the results of our calculations using syntactic complexity metrics.

1.   *Fichtner's C Scores.*

The Fichtner's C index measures the syntactic complexity of verb structures.[239] The higher the Fichtner's C score, the more syntactically complex the text is. Thus, a text with elaborate sentences, prepositional phrases, and subordinate clauses will register a higher score. Figure 2 (below) displays the results of our Fichtner's C calculations including average scores for TOUs, privacy policies, and arbitration clauses within the TOUs. Figure 2 also illustrates scores for the Austen Corpus, the Brown Corpus, and some individual platform TOUs.

---

238.   *See* Rustad & Koenig, *supra* note 4, at 1460–63.

239.   *See supra* notes 179–190 and accompanying text (explaining the Fichtner's C measurement in detail).

Figure 2: Fichtner's C Scores
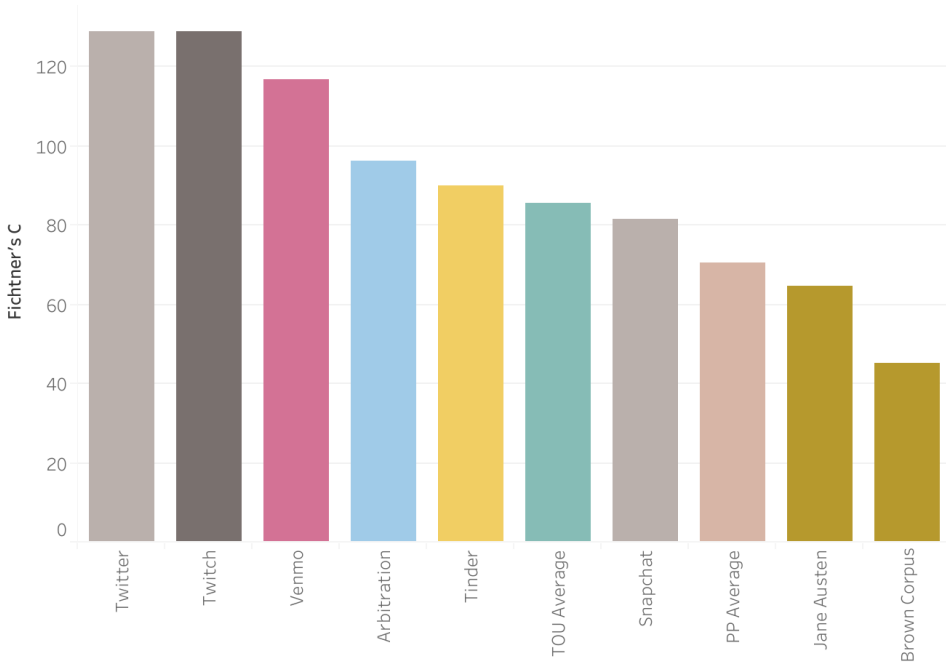
Within TOUs, arbitration clauses produce especially high Fichtner's C scores, scoring even above the average individual TOU.[240] Privacy policies, on the other hand, register less verb complexity. Similarly, the Austen Corpus, which is the most readable according to traditional formulas, has a significantly higher Fichtner's C score than the Brown Corpus. On other metrics—the two traditional readability and MATTR scores—the Austen and Brown corpora tracked more closely. The difference in verb complexity may be related to structural tendencies across text genres.[241] The Austen Corpus, for instance, contains much more character dialogue than the Brown Corpus.[242] Still, the Austen and Brown corpora registered significantly lower C index scores than the TOU average.

---

240. This exposes a potential gap in traditional readability metrics because there is not a direct correlation between these groups of results.

241. For a discussion of syntactic tendencies across genres, see DOUGLAS BIBER & SUSAN CONRAD, REGISTER, GENRE & STYLE (2019). Though the Brown Corpus contains some fiction, it contains a wide range of general prose in modern English. *See supra* notes 214–220 (describing the Austen and Brown corpora).

242. Fiction is just one component of the Brown Corpus. *See supra* notes 214–220 (describing the Austen and Brown corpora).

2.   TAASSC: Noun Phrase Elaboration Scores.

Here, we report the results of NP elaboration, the composite score of all nineteen TAASSC noun phrase types and embedding indices. Like Fichtner's C, the higher the NP score, the greater the syntactic complexity. Figure 3 (below) displays those results across the Jane Austen Corpus, the Brown Corpus, and selected individual platform TOUs.

Figure 3: TAASSC Noun Phrase Elaboration
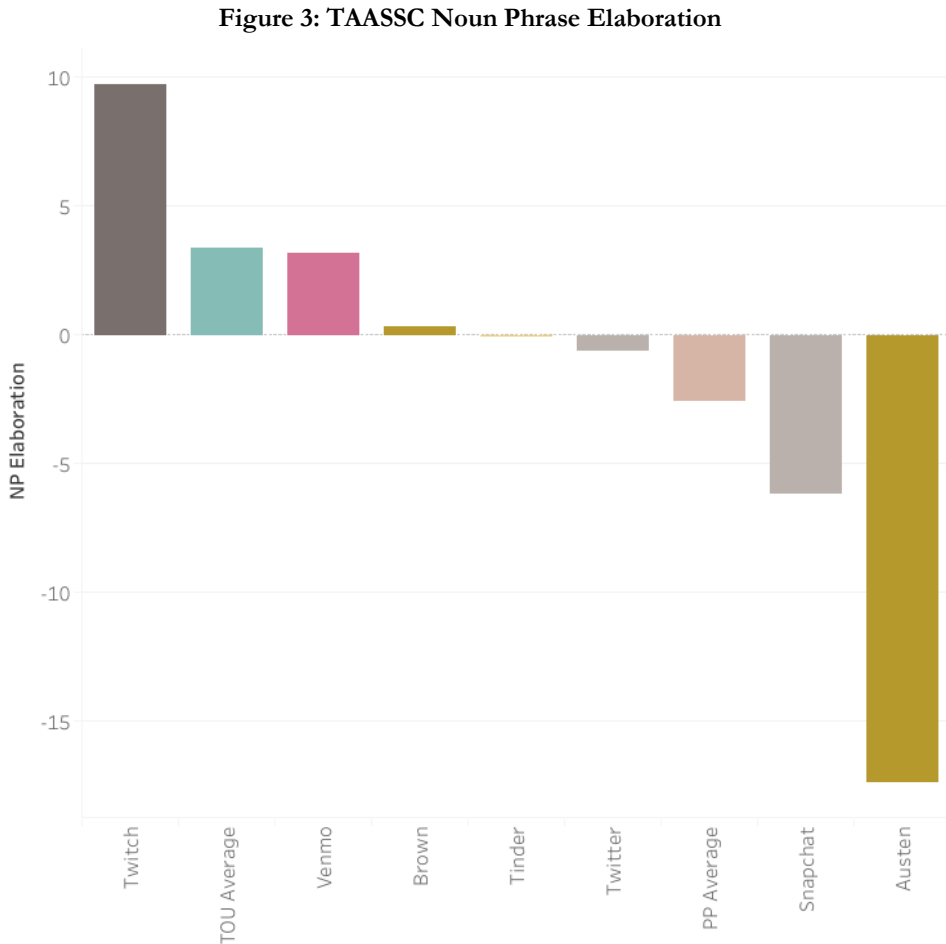
The NP elaboration results indicate that TOUs contain highly complex noun structures, such as embedding. Embedded phrases weigh heavily in the TAASSC results, which is a composite score. Embedding has proven a key

element in reading difficulty.[243] As the TAASSC results suggest, embedding is especially prevalent in academic and legal writing. Consider, for instance, one type of embedding (prepositions per nominal) from the first sentence in the Twitch TOU:

> Welcome to the services operated by Twitch Interactive, Inc. (collectively with its affiliates, "Twitch" or "We") consisting of the website available at https://www.twitch.tv, and its network of websites, software applications, or any other products or services offered by Twitch (the "Twitch Services").

This sentence—though not overwhelming in length—is rather awkward to read. Embedding has a lot to do with that. Consider, for instance, the prepositions per nominal. They present in multiple embedded phrases: *to the services*, *by Twitch Interactiv*e, *with its affiliates*, *of the website*, *at twitch.tv*, *of websites*, and *by Twitch*. On its own, the occasional prepositional phrase is not overly perplexing. However, once several of them are embedded in a single sentence, the text quickly becomes more difficult to read.

Variations in NP scores appear to track with distinctive tendencies across genres and categories. For instance, the gap between the Austen Corpus and TOUs is especially stark in NP complexity. That result indicates that that the literary prose in the Austen Corpus contains far fewer prepositions per nominal and other types of embedding such as determiners and adjectival modifiers. A similar tendency registers within categories of the Brown Corpus. For instance, categories of fiction in the Brown Corpus produce low NP complexity scores—on par with the Austen Corpus. Meanwhile, NP scores are very high within divisions of the Brown Corpus that contain academic and legal texts—on par with the TOU average.[244]

Across both syntactic complexity metrics—Fichtner's C index and the NP index—TOUs register especially high syntactic complexity. Our results suggest a high degree of overall difficulty and sophistication across TOUs.[245] Our results also reveal a curious divergence between privacy policies and TOUs. Whereas traditional readability scores for TOUs and privacy policies are

---

243. Studies have shown that embedding—for instance, center-embedded clauses—is a key factor in reading difficulty. *See supra* notes 196–200 and accompanying text.

244. The "learned" category of the Brown Corpus contains academic, technical, and scholarly texts. Another miscellaneous category contains mostly government and legal texts. Those two categories had even higher NP scores than the TOU average. *See supra* notes 215–216 (describing the contents of the Brown Corpus).

245. Readers may find that observation consistent with examples (provided in Part III, above) of sentences that contribute to higher complexity scores. Those fragments came from the Twitch (for Fichtner's C) and TikTok (for NP elaboration) TOUs, respectively. *See supra* notes 193, 207 and accompanying text.

virtually the same, they diverge in our syntactic complexity results.[246] In verb and noun structures, TOUs score as significantly more complex than privacy policies.[247] Those results align with our anecdotal observations.[248] In sum, traditional readability metrics appear to overlook linguistic differences between privacy policies and TOUs. That divergence underscores doubts about the validity of traditional readability metrics and deserves further research.[249] For now, we speculate that traditional readability formulas might be overlooking linguistic features that make TOUs especially difficult to read.

C.   LEXICAL RICHNESS

The vocabulary found within a text plays a distinctive role in linguistic structures and overall text complexity. Lexical richness refers to the number of unique words used in a text—in other words, the variety of the vocabulary. We illustrate lexical richness with MATTR, a reliable indicator of lexical diversity.[250] For MATTR scores, the scale ranges from zero to one. A higher MATTR score indicates more lexical richness. Figure 4 (below) displays the results of our MATTR calculations for individual TOUs from selected platforms as well as all arbitration clauses in our dataset. Figure 4 also illustrates MATTR results for the Austen and Brown corpora.

---

246.   The median FRE scores of TOUs and privacy policies are almost identical. *See supra* note 237 and accompanying text.

247.   The median Fichtner's C (verb complexity) score for TOUs is 86.3 while the median score for privacy policies is 68.1. The median NP score (noun complexity) for TOUs is 4.38 while the median score for privacy policies is −2.5. *See supra* Figures 2–3.

248.   In our exposure to the texts throughout this study, we found the language in privacy policies generally easier to read than TOUs.

249.   *See supra* notes 172–178.

250.   *See* BREZINA, *supra* note 209, at 58–59; Covington & McFall, *supra* note 23., at 95–96, 99.

Figure 4: MATTR Scores
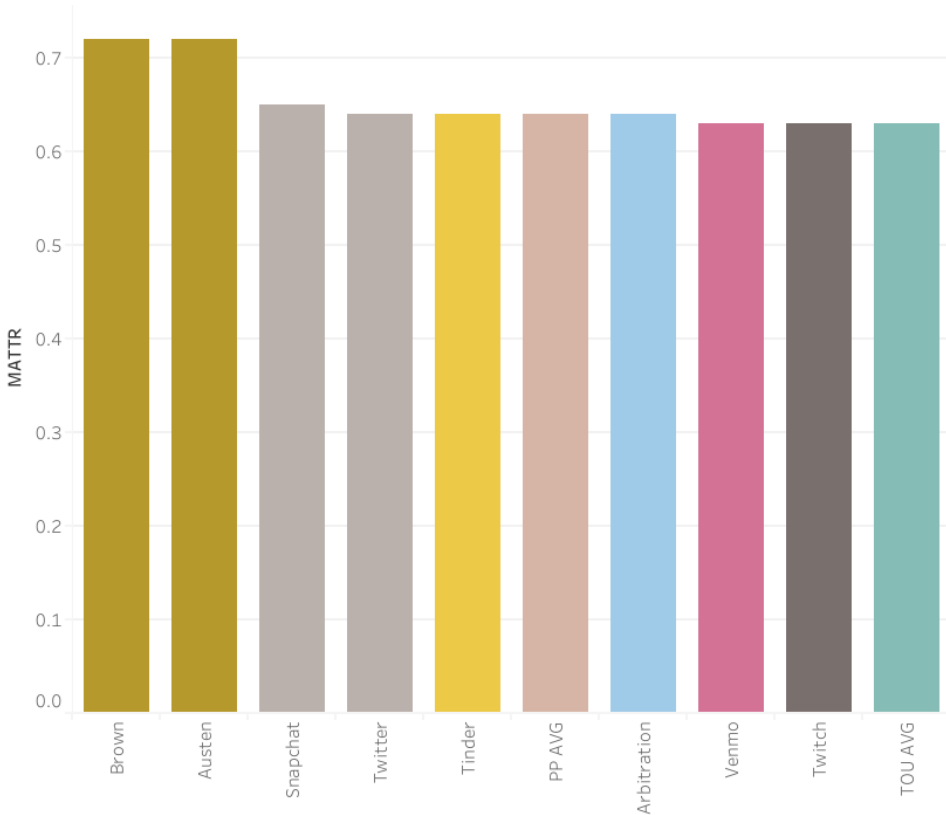
The corpora with the highest readability scores and the lowest linguistic complexity scores also register highest in lexical richness. In a way, that result may seem counterintuitive. We might expect higher degrees of lexical richness—in essence, more diverse vocabularies—to be associated with more challenging and complex texts. However, we observe the opposite. For instance, the Jane Austen and Brown corpora, which have significantly higher MATTR scores, also exhibit lower syntactic complexity and higher FRE scores. In other words, they have more diverse vocabularies yet are also easier to read.

A potential explanation for those correlations: TOUs tend to repeat complex legal jargon, frequently, in long sentences. That tendency produces lower lexical diversity *and* relatively difficult texts. For instance, the word "indemnification," if used frequently enough throughout a text, could simultaneously reduce traditional readability scores and the MATTR score of a text. Arbitration clauses exhibit a similar pattern on the linguistically difficult end of the spectrum: low MATTR scores with high complexity and low

readability scores. A similar explanation likely applies.[251] These patterns suggest that lexical diversity—as opposed to syntactic complexity or perhaps even lexical difficulty—is not an ideal proxy for understanding the linguistic difficulty of contract texts.

D.   METADATA

We gathered metadata for several variables in two divisions: platform-level metadata and TOU-level metadata. The platform-level variables we collected are category, domicile, and user base.[252] Among the TOU-level variables we gathered: word count, dispute resolution and jurisdiction, modification, and governing law. For TOUs that contain arbitration clauses, we also collected data on opt-outs and class waivers. Below we illustrate and discuss some of the results of our metadata analysis.

1.   The "TL" in TL;DR: Word Count

Our dataset exhibits a remarkably wide range in word count. The shortest, Telegram, with just seventy-five words, is something of an outlier.[253] Telegram, founded by Russian entrepreneurs and headquartered in Dubai, exhibits unusual features. In addition to its extremely low total word count, the Telegram TOU is silent on critical questions like governing law and dispute resolution. At the high end of the range is Venmo's TOU at 20,505 words, which is situated around the length of a shorter novella or a law review article.[254] The average length across the TOUs in our dataset is 6,712 words; the median is 5,830 words. Figure 5 (below) illustrates the word counts of individual apps alongside the mean and median word counts of TOUs and

---

251. For instance, a sample sentence from Snapchat's arbitration clause: "Notwithstanding any other provision of this Agreement, the Arbitration Agreement or ADR Services' Rules, disputes regarding the interpretation, applicability, or enforceability of this waiver may be resolved only by a court and not by an arbitrator." *Snap Inc. Terms of Service*, SNAP INC., https://snap.com/en-US/terms (last updated Aug. 15, 2023).

252. We exclude user base from our analysis because data on active users is unreliable. For larger platforms, particularly those that are publicly listed, data is widely available and somewhat reliable. *See supra* note 152 and accompanying text. However, for smaller or unlisted platforms, data on active users is intermittent at best.

253. After our scrape, Telegram added more TOU language in an update when the platform rolled out Telegram Premium in June 2022. *See 700 Million Users and Telegram Premium*, TELEGRAM (June 19, 2022), https://telegram.org/blog/700-million-and-premium. Still, even at 1,412 words, the Telegram TOU remains exceptionally short. *See Terms of Service*, TELEGRAM, https://telegram.org/tos/terms-of-service-for-telegram-premium (last visited Sept. 2, 2022).

254. *See* Chuck Sambuchino, *How Long is a Novella? And How Do You Query Agents For Them?*, WRITER'S DIGEST (Nov. 18, 2008), https://www.writersdigest.com/publishing-insights/how-long-is-a-novella-and-how-do-you-query-agents-for-them ("Novellas generally run 20,000–50,000 words. About 30,000 words are average.").

privacy policies. The TOU range is dramatic with the highest at 20,505 words (Venmo) and the lowest at just 75 words (Telegram).

Figure 5: Word Counts by Agreement Category and Apps.
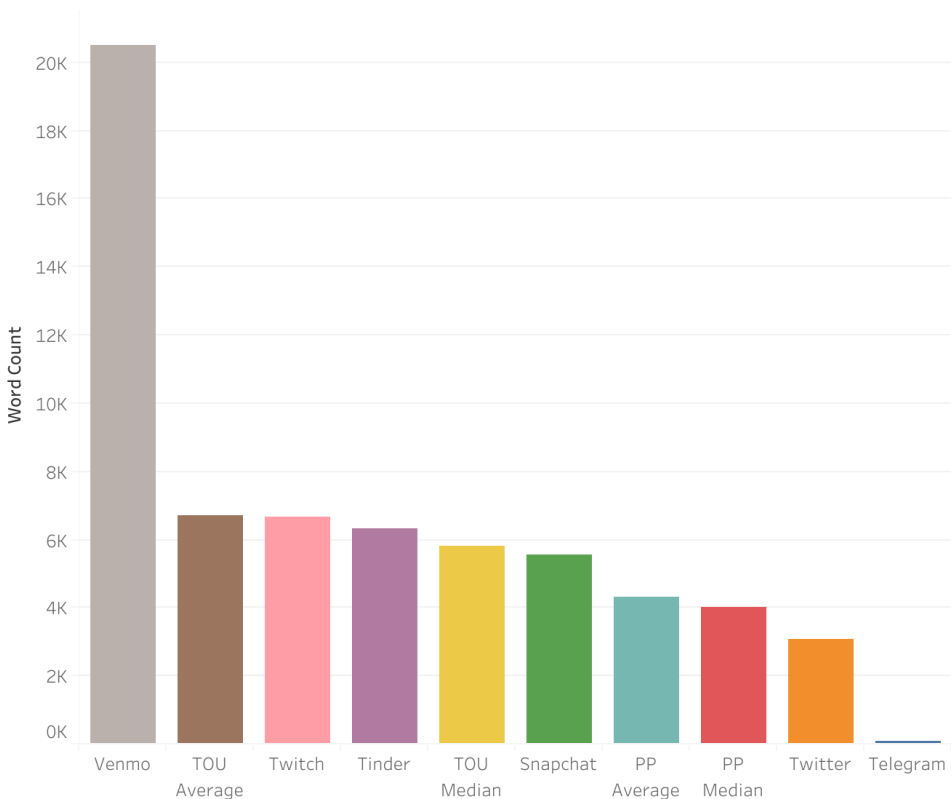
Overall, the volume of TOUs is extraordinary. Word count volumes are substantial and appear to increase over time, as Table 2 (below) illustrates. We observe a meaningful increase in word counts in comparisons with other TOU and privacy policy datasets. For instance, our dataset is substantially similar to the dataset compiled by Rustad and Koenig, as both datasets focus the consumer contracts of social platforms. Whereas our dataset includes 195 primary terms, including seventy-five TOUs, theirs includes 329 TOUs. Across the two datasets with a roughly eight-year time horizon, the median word count jumps from 3,910 words to 5,830 words. Table 2 shows a substantial increase in TOU word counts across these two datasets. This

comparison is consistent with other studies finding that online TOUs and privacy policies have expanded in length over time.[255]

Table 2: Word Counts in Social Platform TOUs, 2014–22

Metric	Rustad & Koenig (2014)	Samples, Ireland, Kraczon (2024) (this Article)
Mean	4,418	6,712
Median	3,910	5,830
Range	249 to 37,239	75 to 20,505

On top of TOUs, privacy policies and community guidelines also present considerable burdens in terms of length and reading costs. The average word count across the privacy policies in our dataset is 4,462; the median is 4,150. For community guidelines, we calculate an average word count of 2,477 and a median of 980. The total word count per platform in our dataset: 12,592 words. Thus, to review just the primary terms for a typical social platform in our dataset, a consumer would need to read a substantial amount of complex language. Any such review typically takes place on a smartphone at the time the user downloads the app and registers an account—an environment that is not conducive for reading important legal materials.

We observe anecdotal evidence that privacy policies, like TOUs, have grown substantially in length over time. Table 3 (below) illustrates the median word counts of privacy policies in three different studies. As a study of seventy-five privacy policies from popular websites in the United States, the dataset developed by Aleecia McDonald and Lorrie Cranor has strong parallels with the privacy policies in our dataset.[256] But the median length of privacy policies in our dataset is 4,150 words, compared to 2,514 words in theirs. As another indication of word count trajectory, Isabel Wagner's longitudinal study of website privacy policies found that the average length has quadrupled since 2000.[257] As shown in Table 3, the increase in median word counts across these datasets suggests that privacy policies have become significantly longer in recent years.

---

255. Wagner, *supra* note 24, at 1 ("We find that the length of the average privacy policy has approximately doubled in the last ten years and quadrupled since 2000.").

256. *See generally* Aleecia M. McDonald & Lorrie Faith Cranor, *The Cost of Reading Privacy Policies*, 4 J.L. & POL'Y INFO. SOC'Y 543 (2008).

257. Wagner, *supra* note 24, at 1.

Table 3: Word Counts of Online Privacy Policies (PPs)

Study	Dataset	Median Word Count
McDonald & Cranor (2008)[258]	PPs of 75 popular websites	2,514
Amos, et al. (2009)[259]	910,546 PPs from 108,499 websites	876
Amos, et al. (2019)[260]		1,522
Samples, Ireland, Kraczon (2024) (this Article)	PPs of 75 social platforms	4,150

Whereas platforms incur very little cost in adding terms to online contracts, snowballing length poses enormous transaction and opportunity costs for consumers. If a consumer decides to read TOUs, the time and effort required to read long, technical texts is considerable. McDonald and Cranor estimated an aggregate opportunity cost of $781 billion in 2008. More startling, perhaps, is how much higher those numbers would be today. Adjusted for inflation, that would be $1.053 trillion in 2022 terms.[261] Also, since 2012, the smartphone contracting ecosystem has expanded significantly.[262] TOUs have also grown longer.[263] The median word count in McDonald and Cranor's dataset was just 2,514 words (versus 4,150 in ours).[264] Also, their cost estimate assumes a reading speed of 250 words per minute.[265] People most likely read TOUs at a substantially slower rate.[266]

---

258. McDonald & Cranor, *supra* note 256, at 544.

259. *See generally* Ryan Amos, Gunes Acar, Eli Lucherini, Mihir Kshirsagar, Arvind Narayanan & Jonathan Mayer, *Privacy Policies Over Time: Curation and Analysis of a Million-Document Dataset* (2021), https://dl.acm.org/doi/10.1145/3442381.3450048 (showing changes in readability scores of privacy policies between 2009 and 2019).

260. *Id.*

261. This calculation uses May 2008 to May 2022 as the time horizon. *See CPI Inflation Calculator*, U.S. BUREAU OF LABOR STATISTICS, https://www.bls.gov/data/inflation_calculator.htm (last visited Jan. 20, 2024).

262. Market research indicates that, on average, people have as many as eighty or ninety apps installed on their phone, but they actively use as few as thirty per month or ten per day. SPOTLIGHT ON CONSUMER APP USAGE: PART 1, APP ANNIE 6–7 (2017), http://files.appannie.com.s3.amazonaws.com/reports/1705_Report_Consumer_App_Usage_EN.pdf; Irfan Ahmad, *60+ Fascinating Smartphone Apps Usage Statistics For 2019*, SOC. MEDIA TODAY (Mar. 23, 2019) https://www.socialmediatoday.com/news/60-fascinating-smartphone-apps-usage-statistics-for-2019-infographic/550990.

263. *See supra* Tables 2–3.

264. *See* McDonald & Cranor, *supra* note 256, at 554.

265. *Id.*

266. *See* Marc Brysbaert, *How Many Words Do We Read Per Minute? A Review and Meta-Analysis of Reading Rate*, 109 J. MEMORY & LANGUAGE 1, 15 (2019) (finding that the reading rate for a sample of difficult texts was 203 words per minute versus 261 for easier texts).

Like our dataset of seventy-five social platforms, McDonald and Cranor assessed seventy-five websites. But privacy policies are just one segment of consumer contracting arrangements. Most platforms present multiple contracts and policies. Almost all use a TOU and a privacy policy, but there are other agreements too. In our dataset, which encompasses the entire contracting framework (TOUs *and* supplementary terms), the average total word count per platform is about 12,644—far more than the McDonald and Cranor dataset, which focuses exclusively on privacy policies.

Practically speaking, reading complex documents on a mobile phone is fairly daunting, even when relatively brief. At over 20,000 words, the Venmo TOU is both complex and long.[267] Not only must a consumer read the TOU to understand the contractual arrangement, but also other supplementary terms.[268] In the case of Venmo, that includes a privacy policy (5,302 words), an acceptable use policy (1,095 words), a consent to receive electronic disclosures (973 words), and others that depend on optional functions. For instance, Venmo users who trade cryptocurrency also agree to the Venmo cryptocurrency TOU, adding another 6,743 words.[269] Thus, a Venmo user who enables crypto trading would need to review, at a minimum, roughly 35,000 words. On top of that, consumers are legally bound to unilateral modifications made by platforms, even when they are carried out in minimally transparent updates.[270]

Consumers are assumed by law to have reviewed TOUs when they download an app and click through the installation prompts.[271] Given the nature of mobile platforms, consumers are likely to interact with TOUs on a smartphone screen. Intuitively, that may seem impractical: to read thousands—or even tens of thousands—of words of dense legal text on a phone screen at the moment the app is downloaded. Research confirms that intuition. Considering the practical and cognitive factors at play, understanding human behaviors in response to long consumer contracts is critical, especially

---

267. As part of highly regulated industries, fintech and payment service platforms are likely to have more extensive, longer terms.

268. *User Agreement*, VENMO, https://venmo.com/legal/us-user-agreement (last updated Oct. 6, 2023) (stating that users "agree to comply with the following additional policies and each of the other agreements" that Venmo posts).

269. *Venmo Cryptocurrency Terms and Conditions*, VENMO, https://venmo.com/legal/crypto-terms (last updated Oct. 23, 2023).

270. *See infra* notes 317–328 and accompanying text (discussing modification clauses and practices among platforms).

271. *See supra* notes 51–52 and accompanying text.

for documents that contain dense and incomprehensible language.[272] Research suggests that reading comprehension on smartphones is relatively low.[273] There are also indications that reading comprehension deteriorates with document length.[274]

2. *Disputes & Jurisdiction*

Arbitration has a long history in the United States, particularly as a mechanism for disputes related to labor-management relations and commercial transactions.[275] With support from key decisions by the Supreme Court, arbitration has expanded into a wide variety of settings, including consumer contracts.[276] Our metadata confirms that social platforms have joined that trend. In our dataset, a 64% majority of the TOUs (or forty-eight of seventy-five) contain arbitration clauses. In some instances, arbitration clauses in our dataset are jurisdiction-specific, applicable only to users in the United States.[277] In the United States, federal law—specifically, the Federal Arbitration Act—and decades of case law have produced a very favorable environment for arbitration.[278] As for TOUs without arbitration clauses, most (twenty-one of twenty-seven) provided for litigation while a handful (six of twenty-seven) were either unspecified or unclear as to dispute resolution.

---

272.   *See, e.g.*, OMRI BEN-SHAHAR & CARL E. SCHNEIDER, MORE THAN YOU WANTED TO KNOW: THE FAILURE OF MANDATED DISCLOSURE 28–29 (2014) (finding that long mandated disclosures create cognitive problems for the reader).

273.   Smartphone reading, compared to reading on a paper medium, produces lower reading comprehension. *See generally* Motoyasu Honma, Yuri Masaoka, Natsuko Iizuka, Sayaka Wada, Sawa Kamimura, Akira Yoshikawa, Rika Moriya, Shotaro Kamijo & Masahiko Izumizaki, *Reading on a Smartphone Affects Sight Generation, Brain Activity, and Comprehension*, 12 SCI. REPORTS 1589 (2022). Research also suggests that smartphone reading produces lower comprehension than reading on a desktop computer. *See* R.I. Singh, M Sumeeth & J. Miller, *Evaluating the Readability of Privacy Policies in Mobile Environments*, 3 INT'L J. OF MOBILE HUMAN COMPUT. INTERACTION 1, 71 (2011).

274.   That appears to be true for both mobile and desktop reading environments. *See* Singh et al., *supra* note 273, at 71 ("An inverse dependency between the length of privacy statement and Cloze test scores shows that readers' comprehension drops with length; a similar conclusion was drawn from the desktop environment.").

275.   *See* Linda J. Demaine & Deborah R. Hensler, *"Volunteering" to Arbitrate Through Predispute Arbitration Clauses: The Average Consumer's Experience*, 67 LAW & CONTEMP. PROBS. 55, 55–56 (2012) (describing the expansion of arbitration in the United States).

276.   *See* Michael L. Rustad, Richard Buckingham, Diane D'Angelo & Katherine Durlacher, *An Empirical Study of Predispute Mandatory Arbitration Clauses in Social Media Terms of Service Agreements*, 34 U. ARK. LITTLE ROCK L. REV. 643, 676–80 (2012).

277.   *See, e.g.*, DISCORD, *supra* note 142 ("IF YOU'RE A U.S. RESIDENT, YOU ALSO AGREE TO THE FOLLOWING MANDATORY ARBITRATION PROVISIONS").

278.   *See, e.g.*, Moses H. Cone Mem'l Hosp. v. Mercury Constr. Corp., 460 U.S. 1, 24 (1983) (framing the FAA as a "congressional declaration of a liberal policy favoring arbitration agreements, notwithstanding any state substantive or procedural policies to the contrary").

Importantly, almost all the arbitration clauses (forty of forty-eight) in our dataset contain class waivers, which prevent a user from participating in collective or class actions of any kind.[279] A typical class waiver, as presented as part of an arbitration clause in the Truth Social TOU:[280]

> . . . YOU AGREE TO ABSOLUTELY AND UNCONDITIONALLY WAIVE ANY AND ALL RIGHTS TO PARTICIPATE IN OR TO BE INCLUDED IN ANY CLASS ACTION LAWSUITS OR INCLUSION IN ANY MULTI-PARTY ACTIONS OR SUITS AGAINST US.[281]

Businesses began including class action waivers—primarily in consumer contracts and employment agreements—to reduce the risk of class action litigation.[282] As these waivers proliferated, so did their controversies. Questions about waivers and class arbitration have appeared frequently before the Supreme Court in recent years.[283] Class waivers are particularly sensitive when they appear in unilaterally drafted agreements with consumers and employees. Courts were initially reluctant to enforce these waivers to funnel employees and consumers towards arbitration.[284] Despite that initial hesitancy, a series of opinions from the Supreme Court reinforced the use of class waivers in a wide variety of contexts, including consumer contracts.[285]

Consumers, particularly residents of the United States, have seen their access to the judicial system dramatically curbed in the online environment. Rights to a jury trial and collective mechanisms are regularly waived in routine

---

279. Particularly controversial in employment and consumer contracts, class waivers have been litigated heavily in recent years, including some prominent SCOTUS cases.

280. *Truth Social Terms of Service*, TRUTH SOCIAL, https://help.truthsocial.com/legal/terms-of-service (last updated Feb. 18, 2022).

281. Another feature of TOUs we reserve for future research is the role of all-caps text in the overall linguistic difficulty for consumers. This question has been addressed in laboratory experiments. *See* Yonathan A. Arbel & Andrew Toler, *ALL-CAPS*, 17 J. EMPIRICAL L. STUD. 862 (2020).

282. *See* JON O. SHIMABUKURO & JENNIFER A. STAMAN, MANDATORY ARBITRATION AND THE FEDERAL ARBITRATION ACT, CONG. RES. SERV. 11–12 (Sept. 20, 2017), https://sgp.fas.org/crs/misc/R44960.pdf.

283. *See* Gary Born & Claudio Salas, *The United States Supreme Court and Class Arbitration: A Tragedy of Errors*, 2012 J. DISP. RESOL. 21, 21 (2012) ("[T]he U.S. Supreme Court has issued a series of confusing and, at times, confused opinions on class arbitration.").

284. *See, e.g.*, Ryan Miller, *Next-Gen Arbitration: An Empirical Study of How Arbitration Agreements in Consumer Form Contracts Have Changed after Concepcion and American Express*, 32 GEO. J. LEGAL ETHICS 793, 794–99 (2019).

285. *See* AT&T Mobility LLC v. Concepcion, 563 U.S. 333 (2011) (overturning a California rule that classified many collective-arbitration waivers in consumer contracts as unconscionable); Shimabukuro & Staman, *supra* note 282, at 11–14 (reviewing high-profile cases on class waivers).

consumer transactions.[286] As the district court in *Selden v. Airbnb* put it, those fundamental rights are foregone "as a condition of simply participating in today's digital economy" through arbitration provisions in TOUs.[287] The practice is widespread enough to play a plot-defining role in an episode of *Silicon Valley*, a comedy series by HBO that portrays tech start-ups in a satirical light.[288] Our results shed light on the extent to which arbitration and class waivers have reshaped those rights in the smartphone contracting environment.

Our results also yield comparisons with previous empirical work on the use of arbitration in consumer contracts. We illustrate the high-level comparison among those results in Table 4, below. In their 2004 article, Linda Demaine and Deborah Hensler found that arbitration clauses were common in consumer contracts (35.4%), but especially prevalent in the financial category (69.2%).[289] A 2008 study of online retail contracts by Ronald Mann and Travis Siebeneicher revealed that less than a tenth had installed arbitration clauses.[290] Finally, and most akin to our dataset, Rustad and Koenig examined the use of pre-dispute arbitration specifically among the TOUs of social platforms, finding that 29% imposed arbitration on users in their 2014 article.[291] Just eight years later, our results reflect a significantly higher rate: 64%.[292]

The comparisons in Table 4 show a remarkably high frequency of arbitration clauses in social platform TOUs. The frequency of arbitration in our dataset is significantly higher than previous TOU studies, including the Rustad and Koenig dataset, which also focused on social platforms.[293]

---

286. *See* Shelley McGill, *Consumer Arbitration Clause Enforcement: A Balanced Legislative Response*, 47 AM. BUS. L.J. 361 (2010) ("An arbitration clause is included in the contract to insulate the corporation from the punishing effects of class actions and not as a serious choice of alternative forum.").

287. Selden v. Airbnb, Inc., No. 16-CV-00933, 2016 WL 6476934, *2 (D.D.C. Nov. 1, 2016), aff'd, 4 F.4th 148 (D.C. Cir. 2021).

288. The show has won two Emmys and many nominations. Silicon Valley, HBO, https://www.hbo.com/silicon-valley (last visited Sept. 14, 2022).

289. Demaine & Hensler, *supra* note 275, at 62.

290. Ronald J. Mann & Travis Siebeneicher, *Just One Click: The Reality of Internet Retail Contracting*, 108 COLUM. L. REV. 984, 999 (2008) ("Perhaps the most surprising finding is that arbitration clauses appear in less than one-tenth of the contracts (only 44 of 500 retailers)."). We speculate that the number would be higher today.

291. Rustad & Koenig, *supra* note 4, at 1469.

292. We acknowledge that these datasets are similar but not exact matches. Also, arbitration displays significant variation across industries, which makes the Rustad and Koenig dataset an especially relevant analog with ours. *See, e.g., supra* note 289 and accompanying text (showing especially high prevalence of arbitration clauses in the financial industry contracts).

293. *See generally* Rustad & Koenig, *supra* note 4.

Table 4: Frequency of Arbitration Clauses in TOUs.

Study	Sample	Arbitration Clauses (%)
Demaine & Hensler (2004)[294]	161 TOUs of various industries	35.4
Mann & Siebeneicher (2008)[295]	500 TOUs of internet retailers	8.8
Rustad & Koenig (2014)[296]	328 TOUs from social platforms	29.0
Samples, Ireland, Kraczon (2024) (this Article)	75 TOUs of social platforms	64.0

3. *Arbitration Opt-Outs*

Another feature of arbitration clauses and access to justice we examine at the metadata level: opt-outs.[297] Usually embedded in arbitration clauses, opt-outs offer users the ability to decline arbitration as the default mechanism for dispute resolution.[298] Opt-outs are fairly common in our dataset: almost half (twenty-two of forty-eight) of the TOUs with arbitration clauses provide some form of opt-out rights. The legal strategy behind the opt-out trend might be understood as a preemptive measure to defeat unconscionability arguments raised by potential plaintiffs.[299] In *Suarez v. Uber*, for instance, the court dismissed procedural unconscionability in light of the plaintiffs' "absolute right" to opt-out of arbitration.[300]

Yet opt-out rights have major limitations. Many expire within a relatively short period—thirty days, for instance.[301] Also, the instructions for opt-out

---

294. Demaine & Hensler, *supra* note 275, at 62 (finding that "fifty-seven of the 161 sampled businesses (35.4%) included arbitration clauses in their consumer contracts").

295. Mann & Siebeneicher, *supra* note 290, at 987.

296. *See generally* Rustad & Koenig, *supra* note 4, at 1469.

297. *See* Farshad Ghodoosi & Monica M. Sharif, *Arbitration Effect*, 60 AM. BUS. L.J. 235, 255 (2023) (discussing consumer tendencies to opt out of arbitration clauses when informed about them).

298. Typically, when a user opts-out of an arbitration agreement, a jurisdiction clause specifying venue/forum applies. *See, e.g., Terms of Service*, KIK, https://www.kik.com/terms-of-service (last updated Aug. 23, 2021) ("To the extent the arbitration requirement does not apply (if ever), you agree that any action at law or in equity for any Dispute shall be filed only in the state and federal courts located in Los Angeles County, California . . .").

299. *See* Ghodoosi & Sharif, *supra* note 297, at 255.

300. *Id.; see also* Suarez v. Uber Techs., Inc., No. 8:16-cv-166-T-30MAP, 2016 WL 2348706, at *4 (M.D. Fla. May 4, 2016).

301. Most platforms allow thirty days. *See, e.g.,* SNAPCHAT, *supra* note 251 ("To opt out, you must notify Snap in writing no later than 30 days after first becoming subject to this

procedures are buried within arbitration clauses. Our results show that these clauses tend to be long and exceptionally complicated.[302] It is unknown, and perhaps doubtful, whether many consumers read and exercise their opt-out rights.[303] Consumers almost never read TOUs at the moment of contract formation (or, for that matter, within thirty or ninety days of that moment). Even if they do read the relevant segment of the arbitration clause within the opt-out period, consumers might not fully appreciate the consequences of waiving their rights to access courts and participate in class actions. A recent study indicates that consumers are unlikely to opt-out of arbitration even when directly presented with the option in a prompt.[304]

Opt-out procedures, practically speaking, create significant transactional friction. Opting out requires fairly sophisticated knowledge and proactive steps by the user. Sometimes procedural burdens are substantial. Venmo's opt-out procedures, for instance, are remarkably cumbersome. In order to opt-out of Venmo's arbitration agreement, a consumer must *mail* a letter—a physical letter, not a "click" on a device or even an email—to a specific address in San Jose, California.[305] Opting out of the arbitration agreement (printing, filling out, and then mailing a form) is considerably more difficult than entering into the contract (a tap as the user opens the app for the first time).[306] Also worth noting: Venmo's cumbersome modifications to the opt-out procedures were embedded in a seemingly routine TOU update, which actually constituted a unilateral and minimally transparent alteration of material terms.[307]

Perhaps the most remarkable fact about Venmo's burdensome opt-out requirements: they are not so usual. Though some opt-outs provide an email option,[308] several others require physical mailing like Venmo.[309] In a July 26,

---

Arbitration Agreement."). Some exceptions: Bubble (31 days), Discord (90 days), and Gettr (5 days).

302.   *See supra* Table 4.

303.   In July 2022, we contacted legal departments at twenty-one platforms to request data on opt-outs. To date, none have supplied data or substantive information about users' opt-out behaviors.

304.   *See* Ghodoosi & Sharif, *supra* note 297, at 255–60.

305.   *See* VENMO, *supra* note 268.

306.   *See* Adam Levitin, *Venmo's Unfair and Abusive Arbitration Opt-Out Provision*, CREDIT SLIPS (Apr. 26, 2022), https://www.creditslips.org/creditslips/2022/04/-venmos-unfair-and-abusive-arbitration-opt-out-provision.html ("What's so ridiculous about requiring a hand-written form to be sent through the mail is that Venmo will surely digitize the form.").

307.   *See id.* (characterizing the modification that contains the opt-out as "unfair and abusive").

308.   Platforms that allow email notice for opt-outs: Badoo, Bumble, Bunch, Coffee Meets Bagel, Discord, Grindr, Hinge, Match, OkCupid, Snapchat, TikTok, and Tinder.

309.   Platforms that require a written/mailed letter for opt-outs: Kik, Gettr, Imo, Instagram, Public, Venmo, WhatsApp, and Wink.

2022 modification, Instagram implemented arbitration with a mail-in procedure for opt-outs.[310] Two platforms—Her, a dating app, and TextFree, a messaging platform—require both![311] There are other quirks as well. Grindr requires an image of a driver's license as part of the opt-out procedures.[312] The Viber TOU contains a passing mention of an opt-out right in the header of the agreement, but no procedure is ever specified.[313] Gettr, which requires a mailed letter, allows users just five business days for submitting an opt-out notice. But the Gettr clause does not specify whether the five-day time limit applies to a postmarked or actual receipt:[314]

> Unless you give us notice of opt-out within five (5) business days of your first use of the Service, addressed to: 3 Columbus Cir, 20th Floor New York, NY 10019, all actions or proceedings . . . shall be submitted to JAMS (www.jamsadr.com) for final and binding arbitration.

Opt-out procedures offer an example of how design principles have the potential to remake TOUs for consumers, perhaps as a more user-friendly experience.[315] In our dataset, opting out of arbitration tends to be far more difficult than forming the contract. For one, to become aware of the opt-out requires reading the contract, which forming the contract does not. In fact, the vast majority of TOUs are agreed upon before the act of reading. (Arguably, TOUs are not even truly intended to be read by consumers.) Second, the procedure itself: agreeing to a TOU (and related policies) often requires a mere click or a swipe. Procedurally, opt-outs require a lot more work, shifting the

---

310. That modification was made after our initial scrape in January 2022. *Terms of Use*, INSTAGRAM, https://help.instagram.com/581066165581870 (last updated July 26, 2022).

311. *Terms of Service*, HER, https://weareher.com/terms (last updated May 25, 2018) ("You must send your opt-out notice to: hello@weareher.com and 1760 Mission Street, San Francisco, CA, 94103."); *Master Terms of Service*, TextFree, https://www.pinger.com/privacy-policy/terms-and-conditions (last updated Dec. 8, 2023).

312. *Terms of Service*, GRINDR, https://www.grindr.com/terms-of-service (last updated Apr. 30, 2023) ("You must email Your legal name, mailing address . . . email address(es) associated with Your Account(s) to which the opt-out applies, and an unaltered digital image of Your valid driver's license . . .").

313. Despite a clear statement that opt-out rights exist at the beginning of the TOU, the arbitration clause is silent on opt-outs. *See Terms of Use*, VIBER, https://www.viber.com/en/terms/viber-terms-use (last updated Apr. 17, 2023) ("YOU HAVE THE RIGHT TO OPT OUT AS DETAILED IN THE ARBITRATION SECTION BELOW.").

314. *Terms of Use*, GETTR, https://gettr.com/terms (last updated May 17, 2023).

315. These theories, often explored in the business-to-business context, may have compelling application in the business-to-consumer environment. *See generally* Gerlinde Berger-Walliser, Thomas D. Barton & Helena Haapio, *From Visualization to Legal Design: A Collaborative and Creative Process*, 54 AM. BUS. L.J. 347 (2017) (articulating the potential for more innovative contract design).

burden of time and effort to the consumer. These practices beg questions that are not limited to opt-out procedures.[316] Is the overall length and difficulty of TOUs an intentional strategy to deter reading and obfuscate unfavorable terms? Are opt-out procedures designed to enable consumer choices, or part of a legal strategy to defeat unconscionability?

4. Modification

As noted above, many platforms use modifications to update and alter their TOUs. Modification clauses set the terms for future modifications or amendments to an agreement. Schmuel Becher and Uri Benoliel use the term "sneak in contracts" to describe TOUs with unilateral and broad modification clauses.[317] Their project includes a detailed examination of multiple variables within modification clauses.[318] At a high level, our primary finding around these clauses: platforms almost always reserve unilateral and unconditional modification rights. Virtually all the platforms in our dataset—almost 95% (seventy-one of seventy-five)—grant themselves unilateral modification rights in their TOUs.[319] These results are consistent with the more comprehensive modification findings by Becher and Benoliel, which indicate that a vast majority (over 98%) of the TOUs in their dataset include unilateral modification rights.[320]

In our dataset, platforms typically reserve unconditional (or nearly unconditional) rights to modify the contract as frequently as needed and without material limitations.[321] Put simply, modification rights are deeply unilateral among social platform TOUs. Tinder's modification clause, for instance, reads:[322]

---

316. Literature on law and strategy has developed useful frameworks for questions like these. *See generally* Justin W. Evans & Anthony L. Gabel, *Legal Entrepreneurship and the Strategic Virtues of Legal Uncertainty*, 57 AM. BUS. L.J. 593 (2020); George J. Siedel & Helena Haapio, *Using Proactive Law for Competitive Advantage*, 47 AM. BUS. L.J. 641 (2010).

317. Their article offers a comprehensive empirical study of modification clauses and three defining elements of sneak in contracts: unilaterality, breadth, and transparency. *See* Shmuel I. Becher & Uri Benoliel, *Sneak in Contracts*, 55 GA. L. REV. 657, 674–75 (2021).

318. *Id.* at 685 (illustrating the "sneakiness" variables in their study).

319. Modification rights in a small minority (4 of 75) TOUs were vaguely bilateral, unclear, or unspecified.

320. *See* Becher & Benoliel, *supra* note 317, at 681 (finding that 98.54% of the modification clauses in their dataset granted the platform unilateral rights to change the TOU).

321. Our anecdotal observation here is consistent with more detailed findings by Becher and Benoliel. *See supra* note 317, at 681–85 (outlining findings).

322. *Terms of Use*, TINDER, https://policies.tinder.com/terms/intl/en (last updated Nov. 19, 2021).

> We may revise this user agreement and any of the policies listed
> above from time to time. The revised version will be effective at the
> time we post it, unless otherwise noted. If our changes reduce your
> rights or increase your responsibilities we will provide notice to you
> of at least 21 days. We reserve the right to amend this agreement at
> any time without notice, subject to applicable law. By continuing to
> use our services after any changes to this user agreement become
> effective, you agree to abide and be bound by those changes. If you
> do not agree with any changes to this user agreement, you may close
> your account.

Consistent with the above example, the critical common denominator among nearly all modification clauses: open-ended and unilateral rights—for the platform—to amend the TOU at will. Most also stipulate that a user's continued use of the platform constitutes an acceptance of any modifications.[323] Notice obligations for modification exhibit more variation, however. Curiously, the Tinder modification clause (above) contains a notice commitment yet also reserves an almost unqualified right to amend without notice. Those terms seem awkward to reconcile. Other modification clauses contain commitments to provide notifications for material changes to the TOU.[324] However, as illustrated in the Truth Social modification clause, consumers often bear the burden of monitoring the updates and modifications to the TOUs:

> It is your responsibility to periodically review these Terms of Service
> to stay informed of updates. You will be subject to and will be
> deemed to have been made aware of and to have accepted, the
> changes in any revised Terms of Service by your continued use of
> the Site and the App after the date such revised Terms of Service are
> posted.

Many platforms frequently modify their TOUs.[325] Because modification rights tend to be so open-ended and unilateral, adverse changes are an ongoing risk for consumers. Those risks are more than theoretical: Venmo, for instance, recently implemented particularly difficult requirements for consumers who wish to opt-out of arbitration.[326] As another example, last year TikTok unilaterally revised its privacy policies to authorize itself to collect

---

323. *See*, *e.g.*, *Terms of Service*, DISCORD, *supra* note 142 ("If you continue to use our services after the changes have taken effect, it means that you agree to the changes.").

324. *See*, *e.g.*, *Terms of Service*, GRINDR, https://www.grindr.com/terms-of-service (last updated Apr. 30, 2023) ("If Grindr determines, in its sole discretion, that the changes We make to this Agreement are material, We will notify You in advance (e.g., within the App or via email).").

325. *See* Marotta-Wurgler & Taylor, *supra* note 22, at 274–75.

326. *See supra* notes 305–309.

"biometric identifiers and biometric information" from user content.[327] After our scrape, as we wrote and prepared this Article, several of the platforms in our dataset modified their TOUs. Instagram, for one, reinstated an arbitration clause.[328]

Unilateral modification at scale has undeniable efficiencies. With billions or millions of parties involved, modifications with real notice, review, and assent could have major transaction costs. Yet, as with other aspects of form contracting at scale, asymmetry prevails: efficiency is transferred to the drafting party while risk and costs are transferred to consumers.

V.   CONCLUSION

Nowadays, almost everyone has a smartphone. People spend a lot of time—often, several hours per day—on those devices. The average smartphone has dozens of apps, many of which harvest enormous quantities of intimate data. The most influential social platforms have grown systemically important, mediating unprecedented swaths of data, human activity, and commerce.[329] Yet, in the United States, the current state of law and policy means that many digital platforms are effectively self-regulated. That status quo elevates the consequence of platform-to-consumer contracts. As a result, the TOUs of the largest platforms are much more than garden-variety consumer contracts; they are de facto frameworks of digital governance. They often determine profound questions facing technology and society.

Meanwhile, the yawning gap between classic contract doctrine and modern contracting is widening. The advent of wrap contracts prompted a reckoning with contracting fundamentals in the 1990s.[330] Then came the Internet Age. Online TOUs and privacy policies brought a slew of new challenges. Now, with contracts forming on a societal scale through mobile devices, the

---

327. *See Privacy Policy*, TIKTOK, https://www.tiktok.com/legal/privacy-policy-us?lang=en (last updated June 2, 2021); *see also* Sarah Perez, *TikTok Just Gave Itself Permission to Collect Biometric Data on U.S. Users, Including 'Faceprints and Voiceprints'*, TECHCRUNCH (June 3, 2021), https://techcrunch.com/2021/06/03/tiktok-just-gave-itself-permission-to-collect-biometric-data-on-u-s-users-including-faceprints-and-voiceprints (summarizing modifications to the TikTok privacy policy around biometric data).

328. *Terms of Use*, INSTAGRAM, https://help.instagram.com/581066165581870 (last updated July 26, 2022). Instagram's TOU as of January 2022 did not contain an arbitration clause. That omission is curious because Instagram's TOU used arbitration clauses as early as 2012. *See* Jonathan Weber & Dan Levine, *Instagram Retreats on Some Service Terms After Backlash*, REUTERS (Dec. 20, 2012), https://www.reuters.com/article/us-usa-instagram-changes-idUSBRE8BK03K20121221.

329. *See supra* Section II.B.

330. *See supra* Section II.A.

smartphone era has introduced new pressures for traditional doctrines. As TOUs overwhelm consumers in their volume and difficulty, they also overwhelm fundamental tenets of contract law. Yet, despite profound changes in the marketplace and modern consumer reality, contract law remains static. Our findings suggest that the disconnect between contract doctrine and consumer reality is wider than ever.

Using interdisciplinary methods, we illustrate key dimensions of that decoupling. As for volume, we demonstrate that the length of platform-to-consumer contracts transfers substantial burdens and costs to users.[331] Additionally, our longitudinal comparisons with other empirical studies suggest that length has expanded in recent years.[332] In terms of linguistic features, our results illustrate the extraordinary complexity of platform-to-consumer contracts across multiple metrics.[333] In effect, most TOUs are beyond the grasp of almost any audience outside of judges and lawyers. Our longitudinal comparisons with previous research highlight the need for further research into the direction of change (e.g., increasing or decreasing complexity) in platform-to-consumer contracts over time.[334]

Our results quantify dramatic asymmetries in platform-to-consumer contracting. Procedural asymmetries—such as volume, costs, and difficulty—have warped the concepts of reasonable notice and meaningful assent. Put another way, the "signal-to-noise" ratio for consumers is more painful than ever.[335] But there are acute substantive asymmetries as well. Our metadata illustrates some of those tendencies: highly unilateral conditions of modification, the frequency of arbitration clauses and class waivers, and onerous opt-out procedures.[336] In sum, our findings offer evidence that TOUs—already long, difficult, and asymmetrical—have become even longer, more difficult, and more asymmetrical. Finally, as for methodology, this Article

---

331. *See supra* Section IV.D.

332. *See supra* Tables 2–3.

333. *See supra* Sections IV.A–B.

334. *See supra* Table 1.

335. *See* James Grimmelmann, *Saving Facebook*, 94 Iowa L. Rev. 1137, 1182 (2009) ("Between the lawyerly caution, the weasel words, the commingling of many standard terms with the occasional surprising one, the legally mandated warnings and disclaimers, and the legalese, most privacy policies have a painfully low signal-to-noise ratio.").

336. *See supra* Section IV.D. As for longitudinal trends, further research is needed, but there are some indications that TOUs are increasingly asymmetrical in substantive terms as well. Arbitration clauses, for instance, are far more common in our dataset than in previous studies. *See supra* Table 4.

also presents a novel approach to using corpus linguistics methods in legal research, an approach we hope to develop further in future work.[337]

VI.   ANNEX 1: INDICES OF NOUN PHRASE COMPLEXITY

Phrase Types	Description	Label	Example
Passive nominal subject	NP that serves as the syntactic subject of a passive clause	nsubj_pass	[Your account]$^{nsubj\_pass}$ was terminated.
Nominal subject	Subject of a (nonpassive) clause that is an NP	nsubj	[You]^{nsubj} are responsible for safeguarding your account.
Nominal complement	Noun or NP that serves as a complement in a copular clause	ncomp	Twitter is [a social platform].^{ncomp}
Dependent Types	*Description*	*Label*	*Example*
Determiners	Quantifiers, articles, and demonstratives	det	Twitter gave you [a]^{det} notice.
Adjectival modifiers	Adjective that modifies an NP or a noun	amod	Tiktok is a [diverse]^{amod} community.
Prepositional phrases	Prepositional phrase that modifies an NP or a noun	prep	You are responsible [for safeguarding your account].^{prep}
Possessives	Possessive pronoun or noun that modifies a noun or NP	poss	[Our]^{poss} community guidelines support individuals.
Nouns as modifiers	Noun or NP that modifies a noun or NP	nn	We do not allow [terrorist]^{nn} organizations on this platform.

Sources: Kyle (2016), Kyle & Crossley (2018).

Notes: This table displays selected examples of noun phase types and dependent types measured in TAASSC Noun Phrase Elaboration. For the sake of clarity, we do not include the full list of indices, which are available at the sources cited directly above. The example sentences are selected and edited text samples from our dataset.

---

337.   *See supra* notes 18–22, 139–141 and accompanying text (distinguishing our methods in this Article from legal corpus linguistics scholarship geared towards judicial interpretation).

VII.   ANNEX 2: SELECTED TOU RESULTS & METADATA

Platform	Category	FRE	MATTR	Fichtner's C	TAASSC	Word Count	Governing Law
Badoo	Dating	39.55	.65	96.07	1.85	9053	Texas
BeReal	Social	22.21	.65	51.48	4.97	2355	France
Bigo Live	Creator	31.28	.64	73.38	7.5	3238	Singapore
Brainly	Q&A	27.81	.61	105.69	7.12	10611	New York
Bumble	Dating	35.85	.64	84.87	5.41	8716	Texas
Bunch Group	Chat	29.23	.63	83.56	3.54	4536	California
Coffee M. Bagel	Dating	29.76	.63	85.01	−1.1	5764	California
Discord	Chat	33.13	.63	97.68	5.84	7135	California
eHarmony	Dating	25.93	.65	67.65	5.59	9633	California
Facebook	Social	43.76	.65	72.91	−16.18	4140	California
Gab	Social	29.31	.63	77.63	11.49	4897	Pennsylvania
Gettr	Social	27.02	.64	112.37	5.29	11197	New York
Grindr	Dating	40.86	.61	66.51	7.53	11282	California
GroupMe	Chat	42.09	.63	47.62	1.22	14449	Other
Hago	Chat	27.67	.64	145.05	6.67	7774	Singapore
happn	Dating	34.6	.65	46.69	10.16	8349	France
Helo	Social	27.04	.63	158.71	7.33	6950	Singapore
Her	Dating	38.05	.62	72.88	1.42	6763	California
Hinge	Dating	29.77	.65	83.56	4.29	12041	Texas
Hoop	Dating	46.20	.63	58.41	−9.47	3180	France
Imgur	Creator	41.25	.66	72.45	−3.68	1818	California
imo	Chat	36.32	.65	86.31	0.74	5095	California
Instagram	Social	42.74	.65	75.29	−12.5	3345	California
Kik	Chat	28.26	.63	96.34	0.09	7514	California
Likee	Creator	28.19	.63	110.46	4.57	4110	Singapore
Line	Chat	43.26	.60	134.18	10.65	3383	Japan
LinkedIn	Social	39.55	.63	69.60	−9.7	4323	California
MarcoPolo	Chat	37.17	.64	71.39	2.33	5702	California
Match	Dating	32.11	.65	100.09	4.38	12282	Texas
MeWe	Social	35.31	.63	74.01	1.16	4787	California
OK	Social	32.49	.59	72.37	23.65	8884	Russia
OkCupid	Dating	26.18	.64	111.53	4.79	12418	Texas
Parler	Social	28.73	.64	95.72	5.3	2964	D.C.
Pinterest	Social	43.61	.67	55.09	−4.83	2303	California
Public	FinTech	26.19	.63	107.03	4.59	9227	New York
QQ	Chat	33.5	.61	86.75	6.48	6787	China
Quora	Q&A	37.53	.64	62.45	4.96	5479	California
Qzone	Social	21.25	.59	125.74	2.99	4706	Hong Kong

Platform	Category	FRE	MATTR	Fichtner's C	TAASSC	Word Count	Governing Law
Rave	Chat	19.13	.61	119.67	5.16	7031	Ontario
Reddit	Subculture	33.41	.63	106.25	−3.78	3765	California
Rumble	Creator	25.17	.63	101.44	9.66	5358	Texas
ShareChat	Chat	40.8	.64	57.66	−3.49	4067	India
Signal	Chat	38.38	.63	43.10	−3.15	1545	California
Sina (Weibo)	Social	33.76	.62	64.05	15.24	3308	California
Skype	Chat	42.09	.63	47.62	1.22	14449	Other
Slack	Chat	21.96	.64	70.43	11.91	5830	California
Snapchat	Chat	39.25	.65	88.02	−6.19	5543	California
Soul	Dating	38.85	.63	109.43	6.22	11916	China
Swarm	Dating	46.27	.62	48.33	−4.14	1666	California
SweetMeet	Dating	30.23	.62	106.82	14.27	5064	Czech Rep.
Tagged	Dating	29.68	.61	95.22	8.7	6292	California
Tango	Creator	28.02	.62	108.37	5.49	12533	California
Tantan	Dating	15.48	.65	144.97	6.99	4109	Hong Kong
Teams	Chat	42.09	.63	47.62	1.22	14449	Other
Telegram	Chat	47.23	.73	15.80	−5.03	75	Other
Textfree	Chat	27.92	.62	117.04	3.16	9744	California
The League	Dating	35.38	.63	96.59	5.68	9336	California
TikTok	Creator	28.78	.63	144.32	6.89	7497	Other
Tinder	Dating	31.95	.64	107.39	−0.06	6338	Texas
Tiya	Chat	34.97	.64	75.43	−0.86	1827	Singapore
Truth Social	Social	24.64	.62	145.89	5.28	7934	Florida
Tumblr	Social	32.49	.64	98.76	1.99	6620	New York
Twitch	Creator	17.8	.63	110.25	9.79	6644	California
Twitter/X	Social	22.15	.64	136.44	−0.62	3074	California
Venmo	FinTech	25.03	.63	141.14	3.2	20505	New York
Viber	Chat	36.23	.63	78.72	0.32	8788	England
Vimeo	Creator	33.48	.67	53.48	−1.89	5012	New York
VK	Social	28.86	.60	83.03	29.04	7387	Russia
Wattpad	Creator	45.87	.66	61.41	−1.85	2804	Ontario
WeChat	Chat	29.11	.60	98.77	4.85	8596	Singapore
Weixin	Chat	25.04	.63	95.03	12.47	5775	China
WhatsApp	Chat	28.78	.63	91.76	0.85	5289	California
Wink	Dating	34.7	.65	99.97	−0.8	9105	California
Youtube	Creator	34.84	.64	82.81	−3.32	3606	California
Yubo	Chat	35.44	.65	68.73	6.9	3954	France
Average	—	32.80	.63	88.75	3.44	6720	—
Median	—	32.49	.63	86.31	4.19	5830	—
Mode	—	42.08	.63	47.62	4.38	14449	California