

# How Smart Are Smart Readers? LLMs and the Future of the No-Reading Problem

*Yonathan A. Arbel\* & Shmuel I. Becher\**

**Abstract.** Large Language Models (LLMs) can be used to summarize and simplify complex texts. In this study, we investigate the extent to which state-of-the-art models can reliably operate as ‘smart readers’: applications that empower consumers to tackle lengthy, difficult-to-read, and inaccessible standard form contracts and privacy policies.

Our analysis reveals that smart readers (1) reduce by 66.9% the length of contracts; (2) reduce reading time by 14:41 minutes (3) improve text readability by converting college-level texts to texts readable by fifth-grade students; and (4) do so without considerably compromising the essential information in the original contracts. Despite these impressive results, smart readers are not flawless. They sometimes miscommunicate legal terminology and occasionally present information in a misleading or erroneous manner. Such issues prevent smart readers from replacing the advice of a qualified lawyer. However, for the large mass of daily transactions where consumers would not consider using a lawyer, current-generation smart readers could be an effective tool. We thus conclude that current generation smart readers have arrived and that their arrival invites an academic and policy paradigm change.

---

\* Associate Professor of Law, University of Alabama, School of Law

\* Professor of Law and Associate Dean (Research), Victoria University of Wellington; Lee Kong Chian Visiting Professor of Law, Yong Pung How School of Law, Singapore Management University. We thank Victoria University of Wellington for financial supports, Tim Samples and the editors of this collections for their feedback on a previous version, and Nicholas Takton for outstanding editorial work.

## TABLE OF CONTENTS

|      |   |    |
|------|---|----|
| I.   | Introduction.....   | 3  |
| II.  | Setting the Scene: LLMs & the No-Reading Problem.....                         | 4  |
| III. | Dataset and Methodology .....   | 8  |
| IV.  | High-Level Results .....  | 12 |
|      | A. <i>Simplification Assessment</i> .....                                     | 12 |
|      | 1. Text Length .....  | 12 |
|      | 2. Text Complexity .....  | 13 |
|      | 3. Text Readability .....   | 14 |
|      | B. <i>Quality Assessment</i> .....  | 16 |
| V.   | Simplification & Quality: Specific Clauses.....                               | 17 |
|      | A. <i>Wall Street Journal: Changes to Subscriber Agreement</i> .....          | 18 |
|      | B. <i>Wall Street Journal: Agreement to Arbitrate</i> .....                   | 20 |
|      | C. <i>Airbnb: Collecting Personal Information from Third Parties</i> .....    | 23 |
|      | D. <i>Netflix: Cancellation</i> .....   | 26 |
|      | E. <i>Amazon: Reviews, Comments, Communications, &amp; Other Content</i> .... | 27 |
|      | F. <i>Amazon: Risk of Loss</i> .....  | 30 |
|      | G. <i>Yahoo: Information Sharing</i> .....                                    | 32 |
|      | H. <i>Spotify: Liability Limitation and Claim Filing</i> .....                | 35 |
| VI.  | Simplification of Specific Clauses: Discussion .....                          | 38 |
| VII. | Summary .....   | 40 |

## TABLE OF FIGURES

|            |   |    |
|------------|---|----|
| Figure 1:  | Length Reduction (in Thousand Words) .....                          | 13 |
| Figure 2:  | Aggregated Reduction Results: Words, Sentences & Reading Time ..... | 13 |
| Figure 3:  | Difficult Words .....   | 14 |
| Figure 4:  | Text Readability Flesch-Kincaid .....                               | 15 |
| Figure 5:  | Text Readability CRM .....  | 16 |
| Figure 6:  | Clause Simplification, WSJ.....                                     | 20 |
| Figure 7:  | Clause Simplification, WSJ (2) .....                                | 23 |
| Figure 8:  | Clause Simplification, Airbnb.....                                  | 25 |
| Figure 9:  | Clause Simplification, Netflix .....                                | 27 |
| Figure 10: | Clause Simplification, Amazon .....                                 | 30 |
| Figure 11: | Clause Simplification, Amazon (2).....                              | 32 |
| Figure 12: | Clause Simplification, Yahoo .....                                  | 35 |
| Figure 13: | Clause Simplification, Spotify.....                                 | 38 |

## I. INTRODUCTION

An organizing problem in consumer contracts is the no-reading problem.<sup>1</sup> The common view in the scholarship is that consumers rarely read standard form contracts,<sup>2</sup> and, therefore, their manifested assent to them is superficial.<sup>3</sup> If consumers indeed do not read (let alone understand) the terms of their transactions, their ability to make informed decisions is doubtful, and sellers' incentive to provide fair and efficient contract terms is undermined.<sup>4</sup>

This chapter evaluates whether *smart readers*—technological tools that use large language models (LLMs) to parse texts—can solve this problem and transform standard form contracting. We evaluate this question by testing current models on their ability to simplify contractual texts. Testing current generation models might seem like writing on ice: the rate of technological improvement is staggering, and whatever results we obtain today will be eclipsed by tomorrow's models.<sup>5</sup> Yet, we engage in this analysis because we want to determine if today's smart readers have already managed to pass a utility threshold. If today's smart readers can empower consumers, this would directly impact the large body of scholarship and policy directed at solving this problem

---

<sup>1</sup> See generally Ian Ayres & Alan Schwartz, *The No-Reading Problem in Consumer Contract Law*, 66 STAN. L. REV. 545 (2014) (describing the no-reading problem and offering a solution to it); see also RESTATEMENT OF CONSUMER CONTS. § 3 Reporters' Notes (AM. L. INST., Tentative Draft 2019) (pointing out that the terms of standard form contracts are “invisible to most consumers” and discussing how firms that modify their contracts must give consumers reasonable notice); Melvin Aron Eisenberg, *Text Anxiety*, 59 S. CAL. L. REV. 305 (1986) (theorizing that when consumers confront the dense text of form contracts, they respond by refusing to read it).

<sup>2</sup> The exact scope of the problem is somewhat contested, but there is at least one domain definitely afflicted by extremely low levels of readership: online end-user license agreements. See, e.g., 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, 22 (2014) (finding that consumers rarely read the terms of end-user license agreements).

<sup>3</sup> Victoria C. Plaut & Robert P. Bartlett, *Blind Consent? A Social Psychological Investigation of Non-Readership of Click-Through Agreements*, 36 L. & HUM. BEHAV. 293, 293 (2012) (noting the “documented phenomenon” of “blind consent” — accepting the terms without reading them — associated with “standard, paper-based contracts” now occurs with online Click-Through Agreements). To be sure, the no-reading problem presents additional challenges, aside from consent, to consumer contracting, and consumers' consent should be doubted for reasons other than the no-reading problem. See, e.g., Michael I. Meyerson, *The Reunification of Contract Law: The Objective Theory of Consumer Form Contracts*, 47 U. MIAMI L. REV. 1263 (1993) (discussing contractual issues in battle of the forms cases where seller includes a disclaimer of warranty of merchantability that the buyer does not read and highlighting the importance of informing consumers, even if they do not read or grasp the terms); Margaret Jane Radin, *Boilerplate Today: The Rise of Modularity and The Waning of Consent*, 104 MICH. L. REV. 1223 (2006) (discussing how even if companies make terms easier to read, consumers will not necessarily read them and asserting that consent is fictional, when, for instance, the terms are filed somewhere inaccessible, as in airline tariffs). Recent work argues that AI contracting technologies, namely “nano contracts,” will autonomously negotiate contracts and circumvent the standard negotiation process and its attendant issues. E.g., Yonathan A. Arbel, *On the Scales of Private Law: Nano Contracts*, 37 HARV. J. L. & TECH. (forthcoming 2024).

<sup>4</sup> For a skeptical view, see Douglas G. Baird, *The Boilerplate Puzzle*, 104 MICH. L. REV. 933 (2006).

<sup>5</sup> On the rapid rise in AI capabilities on a variety of tasks, see Yonathan A. Arbel, Matthew Tokson, Albert Lin, *Systemic Regulation of Artificial Intelligence*, 56 ARIZ. ST. L. REV. (forthcoming 2024).

through non-technological means. If they cannot, however, regulation may be justified in relying on non-technological tools.

To frame our analysis, we offer a brief background on smart readers and their relevance to the no-reading problem in Section II. In Section III, we describe our dataset and methodology. We present the results of our examination at the level of the entire agreement, comparing the complexity, length, readability, and quality of the legal documents before and after their simplification in Section IV. Then, in Section V, we shift the focus from the entire legal text to (eight) specific clauses, allowing for a more in-depth and digestible analysis of the models' capabilities, advantages, and limitations.

In Section VI, we discuss the key insights of this study. We find that smart readers perform well on both quantitative and qualitative metrics. They cut in half text difficulty, shorten long texts considerably, and generally capture the most important or intricate aspects of the original texts they simplify. Yet, smart readers also struggle with some types of clauses and sometimes understate, omit, or provide incorrect information on some contractual aspects. In all, smart readers do not replace the careful eye of an experienced lawyer, but they can address consumer problems at scale, cheaply, efficiently, and effectively. In other words, we find that smart readers have arrived.

## II. SETTING THE SCENE: LLMs & THE NO-READING PROBLEM

While most scholars believe that consumers do not read form contracts and privacy policies (the “no-reading problem”), the *reason* for this phenomenon is not quite settled. Why do consumers abstain from reading? Scholars have offered several explanations. Some focus on rational apathy, with not reading emerging as a rational strategy considering the immediate and real costs of reading against the uncertain future gains of doing so. Consumers may also misperceive contract terms or ignore them altogether if they are prone to myopia, information overload, or other forms of behavioral biases.<sup>6</sup> The take-it-or-leave-it nature of most form contracts also makes reading unattractive for negotiation purposes.<sup>7</sup> Other explanations for consumers' tendency to not read form contracts relate to

---

<sup>6</sup> See OREN BAR-GILL, *SEDUCTION BY CONTRACT: LAW, ECONOMICS, AND PSYCHOLOGY IN CONSUMER MARKETS* (Oxford Univ. Press 2012).

<sup>7</sup> See Nat'l Lab. Rels. Bd. v. Gen. Elec., 418 F.2d 716, 768 (2d Cir. 1969), *cert. denied*, 397 U.S. 965 (1970) (characterizing a “take-or-leave-it” approach as a “hard position” that “may be unattractive”).

reputational constraints, trust and social norms, and a (sometimes misguided) belief in the courts' reluctance to enforce unreasonable terms.<sup>8</sup>

However, perhaps the most influential accounts relate to the writing itself. Consumer form contracts are cognitively taxing, visually difficult, and replete with blocks of off-putting ALL-CAPS while employing arcane terms, complex language, and difficult concepts.<sup>9</sup> Consumers do not read contracts, in short, because reading them is a miserable experience.<sup>10</sup>

These challenges lead to a central problem in unregulated markets. Namely, if consumers do not read forms and the law generally allows them to proliferate, firms can insert self-serving terms without losing demand.<sup>11</sup> This situation gives rise to a winning strategy we dub HIDE. Under HIDE, firms adopt forms that are “Hardly Interpretable but Dependably Enforceable.” The HIDE strategy allows firms to benefit from both worlds: maximizing their share of the transactional surplus while reaping the benefits of legal enforcement.<sup>12</sup>

To deal with HIDE and increase the legitimacy of consumer form contracts, scholars, regulators, courts, and advocates have sought solutions to make contracts more readable and accessible and consumers' assent less questionable. Courts, in a perhaps naïve attempt to incentivize consumers to read, often impose a so-called duty to read.<sup>13</sup> At the same time, lawmakers around the country have instituted hundreds of plain language laws.<sup>14</sup> The UCC famously conditions enforcement of warranty disclaimers on their formatting and

---

<sup>8</sup> The literature here is vast. See, e.g., Yonathan A. Arbel & Roy Shapira, *Theory of the Nudnik: The Future of Consumer Activism and What We Can Do to Stop It*, 73 VAND. L. REV. 929 (2020) (consumer expectations and reputational constraints); Shmuel I. Becher & Tal Z. Zarsky, *Minding the Gap*, 51 CONN. L. REV. 69 (2019) (reputation, conduct, and trust); see also Oren Bar-Gill, *Seduction by Plastic*, 98 NW. U. L. REV. 1373 (2004) (cognitive biases); Shmuel I. Becher, *Behavioral Science and Consumer Standard Form Contracts*, 68 LA. L. REV. 117 (2007) (behavioral phenomena); Robert A. Hillman & Jeffrey J. Rachlinski, *Standard-Form Contracting in the Electronic Age*, 77 N.Y.U. L. REV. 429 (2002) (discussing, among other things, trust and social norms as impediments to reading).

<sup>9</sup> See Yonathan A. Arbel & Andrew Toler, *ALL-CAPS*, 17 J. EMPIRICAL LEGAL STUD. 862, 865 (2020) (using all-caps does not “improve consumer consent in any appreciable manner”); see also Uri Benoliel & Shmuel I. Becher, *The Duty to Read the Unreadable*, 60 B.C. L. REV. 2255 (2019) ((un)readability); Tim Samples, Katherine Ireland & Caroline Kraczon, *TL; DR: The Law and Linguistics of Social Platform Terms-of-Use*, 39 BERKELEY TECH. L. J. (forthcoming 2024) (length).

<sup>10</sup> Eisenberg, *supra* note 1, at 310.

<sup>11</sup> See Meyerson, *supra* note 3, at 1312.

<sup>12</sup> See NANCY KIM, *WRAP CONTRACTS: FOUNDATIONS AND RAMIFICATIONS* 76-87 (Oxford Univ. Press 2013) (exploring the utilization of such terms and the courts' enforcement).

<sup>13</sup> E.g., *Mut. of Omaha Ins. Co. v. Driskell*, 293 So.3d 261, 264 (Miss. 2020) (noting that the insured had “an affirmative duty to read” the insurance policy); see also JOSEPH M. PERILLO, *CALAMARI AND PERILLO ON CONTRACTS* 342 (6th ed., West 2009); John C. Calamari, *Duty to Read – A Changing Concept*, 43 FORDHAM L. REV. 341 (1974) (examining the idea in detail).

<sup>14</sup> Michael Blasié, *Regulating Plain Language*, 2023 WIS. L. REV. 687, 687 (2023) (noting that “legislators and regulators” have “passed over seven hundred plain language laws”).

presentation;<sup>15</sup> the Magnuson-Moss Act<sup>16</sup> controls language in disclaimers; the Truth in Lending Act (TILA)<sup>17</sup> controls both presentation and language. Agency action is also involved. For instance, the CFPB recently published new guidelines on what counts as abusive behavior, which includes ‘buried disclosures’ broadly understood to include ‘the use of fine print, or complex language’.<sup>18</sup>

Such regulations cover broad markets and are quite influential. A prominent example comes from insurance markets. Here, some states and government agencies have adopted not only plain language requirements but also required specific scores on reading metrics, such as the Flesch-Kincaid readability metric.<sup>19</sup>

While well-intentioned, these policies have an uneasy fit to those anonymous mass transactions that typify the consumer experience. Consumers are diverse, and their cognitive and linguistic skills, education, socioeconomic status, life experience, expectations, and visual acuity can differ significantly. Millions of American adults struggle with literacy, for diverse reasons.<sup>20</sup> The idea of plain language homogenizes consumers, supposing that a single serving of simpler words can address the needs of a diverse group. But, in reality, such reforms come at a cost and do not necessarily help those who need protection most. At the same time, writing legal texts that would be readable by those with low literacy is a challenge to the best of writers. Furthermore, plain language efforts often do not address the issue of length: consumers are likely to avoid reading plain language contracts if their length is excessive, and few regulatory frameworks limit the length of legal texts.

Enter smart readers. In 2021, we noticed that the emergent transformer technology shows real promise in processing text in general and legal texts in particular. GPT-2, and later GPT-3, could interact with natural language in ways not conceivable before. The feasibility of developing smart readers— that is,

---

<sup>15</sup> U.C.C. § 2-316.

<sup>16</sup> 15 U.S.C. § 2301 et seq.

<sup>17</sup> 15 U.S.C. § 1601.

<sup>18</sup> CONSUMER FIN. PROT. BUREAU, POLICY STATEMENT ON ABUSIVE ACTS OR PRACTICES 5 (2023), <https://www.consumerfinance.gov/compliance/supervisory-guidance/policy-statement-on-abusiveness/#71> [<https://perma.cc/R6L3-SFB9>]; see also Yehuda Adar & Shmuel I. Becher, *Ending the License to Exploit: Administrative Oversight of Consumer Contracts*, 62 B.C. L. REV. 2405 (2021) (proposing a dynamic preventive model of administrative oversight over consumer contracts).

<sup>19</sup> For a few examples see Benoliel & Becher, *supra* note 9, at 2273-74.

<sup>20</sup> 8.4 million Americans are estimated to be below level 1 on the international PIAAC test, which is considered functionally illiterate; another 8 million are also suspected of falling into this category, although the evidence on this is weaker. Saida Mamedova & Emily Pawlowski, *Adult Literacy in the United States*, NAT'L CTR. FOR EDUC. STAT. (July 2019), <https://nces.ed.gov/pubs2019/2019179/index.asp> [<https://perma.cc/83X4-HRTG>]. For a skeptical view of literacy statistics, see Yonathan A. Arbel, *The Readability of Contracts: Big Data Analysis* (2023) (working paper, on file).

advanced large language models (LLMs) capable of parsing, personalizing, and clarifying legal texts for consumers—is becoming manifestly clear.<sup>21</sup>

As we demonstrated elsewhere,<sup>22</sup> these capabilities mean that, for the first time, readers could have contracts presented in a way that *they* could understand. Instead of serving an abstract *average* or *reasonable* consumer, smart readers could tailor the text to specific, ad-hoc, personalized, or idiosyncratic needs of the individual user. Most promising, the technology was almost entirely consumer-sided. The seller was not part of the process, and any HIDE strategy they might pursue was now challenged. Consumers could have control. Smart readers could penetrate the dense language thicket; each contract could be tailored to the individual consumer.

To be sure, the technology in 2021 was nascent. The models we used were quite clunky and success was sporadic.<sup>23</sup> To showcase its potential, we had to cherry-pick examples, a fact we explicitly noted.<sup>24</sup> GPT-3 produced outputs that were sometimes unreliable and misleading, while other times they were meandering and irrelevant.<sup>25</sup> Understandably, when we presented our work, commentators were often skeptical. One reason for their skepticism was that the technology's inconsistency meant consumers cannot reliably trust it. There were also understandable concerns about the ability of this technology to separate the wheat from the chaff, work at scale, parse complex texts, account for specific legal knowledge, and avoid capture by sellers.

We could not provide hard proof that these issues were temporary. However, the arc of this technology was clear to those immersed in the technical details of how it worked. The problems salient back then were true issues, but they related to insufficient data and compute resources,<sup>26</sup> rather than a missing intellectual breakthrough. At the fundamental level, it was clear that these issues were transient.

At the time of writing this manuscript, the latest state-of-the-art model (GPT-4) has moved from previous generations' worse-than-guesswork on the MBE

---

<sup>21</sup> Yonathan A. Arbel & Shmuel I. Becher, *Contracts in the Age of Smart Readers*, 90 GEO. WASH. L. REV. 83, 111 (2022).

<sup>22</sup> *Id.* at 87.

<sup>23</sup> *Id.* at 89.

<sup>24</sup> *Id.*

<sup>25</sup> *See id.* at 120.

<sup>26</sup> Compute is a term of art referring to a measure of computer resources used for processing information. *See* Lennart, *What is Compute? – Transformative AI and Compute [1/4]*, EFFECTIVE ALTRUISM F. (Sept. 23, 2021), <https://forum.effectivealtruism.org/posts/BHPxe8YuuJ4SZWAF3/what-is-compute-transformative-ai-and-compute-1-4> [<https://perma.cc/2VFJ-QEE8>].

exam to passing the bar<sup>27</sup> and, in fact, surpassing the level of the average test taker. It passed other bars as well. GPT-4 has blazed through the LSAT (88th percentile) and medical exams (75th percentile).<sup>28</sup> It achieved good scores on challenging economics and advanced quantum physics exams.<sup>29</sup> Most importantly, the sentiment changed. While the technology is imperfect in many ways, it became apparent that its current achievements do not represent all of its potential. Indeed, the main question commentators and the public ask today is not what the technology can do, but what it *cannot*.<sup>30</sup>

The rate of mass adoption of ChatGPT has surpassed almost any other technology or invention.<sup>31</sup> Versions of LLMs are now accessible to the end user free of charge. The technology requires little expertise to use. Most of all, the technology is impressive. From a law and policy perspective, the time is ripe to evaluate whether smart readers could empower consumers and solve the no-reading problem.<sup>32</sup>

### III. DATASET AND METHODOLOGY

**Dataset.** Our dataset consists of eight contracts and privacy policies, from diverse key industries, with varying degrees of length and complexity. We selected agreements and policies from some of the most popular businesses and service providers. These documents include (1) Yahoo Privacy Policy, (2) Wall Street Journal Terms of Service, (3) Spotify Terms and Conditions, (4) Snapchat

---

<sup>27</sup> Pablo Arrondo, *GPT-4 Passes the Bar Exam: What That Means for Artificial Intelligence Tools in the Legal Profession*, STAN. L. SCH. (Apr. 19, 2023), <https://law.stanford.edu/2023/04/19/gpt-4-passes-the-bar-exam-what-that-means-for-artificial-intelligence-tools-in-the-legal-industry/> [<https://perma.cc/DE9U-FHGN>].

<sup>28</sup> John Koetsier, *GPT-4 Beats 90% of Lawyers Trying to Pass the Bar*, FORBES (Mar. 14, 2023), <https://www.forbes.com/sites/johnkoetsier/2023/03/14/gpt-4-beats-90-of-lawyers-trying-to-pass-the-bar/?sh=77c790ca3027> [<https://perma.cc/ZTF4-6NSJ>].

<sup>29</sup> Bryan Caplan, *GPT Retakes My Midterm and Gets an A*, BET ON IT (Mar. 21, 2023), <https://betonit.substack.com/p/gpt-retakes-my-midterm-and-gets-an> [<https://perma.cc/LMG4-E3TY>] (economics); Matt Swayne, *ChatGPT-4 Receives 'B' on Scott Aaronson's Quantum Information Science Final — Immediately Emails the Dean Seeking a Better Grade*, QUANTUM INSIDER (Apr. 13, 2023), <https://thequantuminsider.com/2023/04/13/chatgpt-4-receives-b-on-scott-aaronsons-quantum-information-science-final-immediately-emails-the-dean-seeking-a-better-grade> [<https://perma.cc/M8WT-X65A>] (quantum physics).

<sup>30</sup> E.g., Pranshu Verma & Gerrit De Vynck, *ChatGPT Took Their Jobs. Now They Walk Dogs and Fix Air Conditioners*, WASH. POST (June 2, 2023), <https://www.washingtonpost.com/technology/2023/06/02/ai-taking-jobs/> [<https://perma.cc/TD5K-JFG2>]; Meghan Bartels, *You Can Probably Beat ChatGPT at These Math Brainteasers. Here's Why*, SCI. AM. (May 25, 2023), <https://www.scientificamerican.com/article/you-can-probably-beat-chatgpt-at-these-math-brainteasers-heres-why/> [<https://perma.cc/XH95-FAM7>].

<sup>31</sup> Krystal Hu, *ChatGPT Sets Record for Fastest-Growing User Base - Analyst Note*, REUTERS (Feb. 2, 2023), <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/> [<https://perma.cc/B5LD-8MEX>].

<sup>32</sup> Our analysis joins other recent work at the law and technology frontiers. For instance, adopting a different helpful measure, Noam Kolt created a dataset of questions on the content of contracts to test the performance of LLMs as a tool to answer content-related questions. Kolt's work showed that the older generation, GPT-3, could already achieve a 77% precision. Noam Kolt, *Predicting Consumer Contracts*, 37 BERKELEY TECH. L. J. 71, 104 (2022).

Terms of Service, (5) Netflix Terms and Conditions, (6) Google Terms of Service, (7) Amazon Conditions of Use, and (8) Airbnb Privacy Policy.<sup>33</sup>

**Assessed Criteria: Readability, Length and Quality.** Our examination focuses on three key criteria at the heart of the no-reading problem. First is text readability. As noted, many suspect that unreadability deters consumers from reading. Therefore, we sought to examine whether language models can make consumer form contracts and privacy policies more readable.

The most famous readability measure is the Flesch Ease of Reading test, which assesses text readability on a 0-100 scale.<sup>34</sup> This test was later amended to convert the scores to a grade-level equivalent, resulting in the Flesch-Kincaid variant.<sup>35</sup> Microsoft users might be familiar with these readability tests, which are embedded in Word.<sup>36</sup> This test is joined by a battery of other tests: Gunning-Fog, SMOG, Linsear-Write, Automated Readability, and Dale-Chall.<sup>37</sup>

The common ground shared by these tests is that they abstract from the meaning strata of the text and evaluate it based on syntactic features, most commonly sentence length, word syllabicity, and word rarity. These measures are widely used but have been recently critiqued for their limited reliability and validity.<sup>38</sup> One special concern is that these tests are highly manipulable. By choosing a different implementation of the Flesch-Kincaid test one could obtain results that show the *same* text requires 4.6 *extra* years of schooling.<sup>39</sup> To limit

<sup>33</sup> Because the terms and privacy policies referenced here were retrieved with New Zealand IP address, they have been posted on the author's own website for permanency. *Welcome to the Yahoo Privacy Policy*, YAHOO!, <https://battleoftheforms.com/wp-content/uploads/2024/01/Yahoo-Privacy-Policy.txt> (last updated Apr. 2022) [<https://perma.cc/4HC8-SDWE>] [hereinafter *Yahoo! Privacy Policy*]; *Subscriber Agreement and Terms of Use*, WALL ST. J., <https://battleoftheforms.com/wp-content/uploads/2024/01/WSJ-terms.txt> (last updated June 27, 2018) [<https://perma.cc/8KEV-2CBH>] [hereinafter *WSJ Terms*]; *Spotify Terms of Use*, SPOTIFY, <https://battleoftheforms.com/wp-content/uploads/2024/01/spotify-TsCs.txt> (last updated Sept. 14, 2019) [<https://perma.cc/8X8Z-7ED4>] [hereinafter *Spotify Terms*]; *Snap Inc. Terms of Service*, SNAP, <https://battleoftheforms.com/wp-content/uploads/2024/01/Snapchat-terms-of-service.txt> (last updated Nov. 15, 2019) [<https://perma.cc/EP6R-T6SH>]; *Netflix Terms of Use*, NETFLIX, <https://battleoftheforms.com/wp-content/uploads/2024/01/Netflix-Ts-Cs.pdf> (last updated Jan. 5, 2023) [<https://perma.cc/W55H-VAL7>] [hereinafter *Netflix Terms*]; *Google Terms of Service*, GOOGLE, [https://battleoftheforms.com/wp-content/uploads/2024/01/google\\_terms\\_of\\_service\\_en\\_NZ.txt](https://battleoftheforms.com/wp-content/uploads/2024/01/google_terms_of_service_en_NZ.txt) [<https://perma.cc/7BPC-AV3S>]; *Conditions of Use*, AMAZON, <https://battleoftheforms.com/wp-content/uploads/2024/01/Amazon-Conditions-of-Use.txt> (last updated Sept. 14, 2022) [<https://perma.cc/8JGR-3C8Z>] [*Amazon Terms*]; *Privacy Policy for the United States*, AIRBNB, <https://battleoftheforms.com/wp-content/uploads/2024/01/Airbnb-privacy.txt> (last updated Jan. 25, 2023) [<https://perma.cc/J2VT-3PXG>] [hereinafter *Airbnb Privacy Policy*].

<sup>34</sup> John Garger, *Determine the Readability Using the Flesch Reading Ease*, JOHN GARGER (Jan 29, 2020), <https://www.johngarger.com/blog/determine-readability-using-the-flesch-reading-ease> [<https://perma.cc/6TMU-78Z9>].

<sup>35</sup> *Id.*

<sup>36</sup> For one explanation, see Benoliel & Becher, *supra* note 9, at 2273.

<sup>37</sup> Common Education Data Standards, *Assessment Item Text Complexity System*, DEP'T OF EDUC., <https://ceds.ed.gov/element/000907> (last visited July 23, 2023) [<https://perma.cc/5WPV-G846>].

<sup>38</sup> Arbel, *supra* note 20.

<sup>39</sup> *Id.*

these issues, we use the Combined Readability Measure (CRM), which measure averages within implementations of the same test and across a number of readability measures. This approach smooths test difference and produces results that are, at the very least, far less manipulable.

The second criterion we assess is text length. Consumer contracts and privacy policies grow longer over time, and there is reason to believe that lengthy texts dissuade consumers from reading. Accordingly, we analyzed the language models' ability to shorten the sampled eight legal documents.

Finally, even if language models can shorten and simplify legal texts, there is still the concern that this will come at the expense of meaning and context. In other words, simplifying a text can result in losing key facts, important details, nuances, and context. Therefore, our third criterion is text quality. Here, we sought to evaluate to what extent the simplified summaries captured the important legal aspects, risks, obligations, and rights.

**Simplification tools: challenges, selection, and programming.** While some services offer AI-powered summarization, none specialize in contracts. Aiming to fully understand and control the summary process presented three technical challenges. We discuss them in turn.

First, many language models are available to select from and each has different limitations on context size and a different mode of interaction (aka Application Programming Interfaces (APIs)). We addressed this issue by selecting the best models that could be used inexpensively. We believe this simulates well the future direction of smart readers, where they will not necessarily rely on state-of-the-art technology to contain costs. The models we picked, however, were all competent models by today's standards, and, hence, offer a good representation of current capabilities. These models came from two firms: Anthropic (Claude) and OpenAI (ChatGPT).

The second challenge we encountered was that current models have constraining limits on input length, meaning that we could not process the entire contract at once. We handled this challenge by developing our own smart reader. In essence, the code we developed performs the following tasks:

- (1) Handle the communication mode with the various models (API);
- (2) Code for each model its limitations on input length;

- (3) Divide the text into ‘chunks’ that fit the input length limitations;<sup>40</sup>
- (4) Define in code our prompt for the task;
- (5) Use a new Python library called ‘langchain’;<sup>41</sup> the code iteratively asks the models for simplification of the relevant chunks; and
- (6) Use the code to combine the chunks together into the resulting simplified contract.

The third challenge involved devising a specific prompt that would ensure our goals as described above: (1) no loss of information, (2) simpler language, and (3) shorter language. Devising this prompt was arguably the most critical aspect in our design since the wording of the prompt can radically change the quality of the output generated by the model. But because there are no robust prompt optimization algorithms that we could use, we relied instead on trial and error.

After some experimentation, we decided on the following prompt:

Simplify contract, low Flesch Kincaid score, KEEP MEANING. Use short words, not legal terms. Swap: accomplishment=success, responsibility=duty, extravagant=fancy. Keep necessary legal concepts. Short sentences. Preserve legal aspects. NO COMMENTS.

It is worth noting three points regarding the prompt. First, the prompt was shorter than we wanted, but it was necessary to keep it short to limit the burden on the maximum input length. It is possible that a more elaborate prompt would yield better outcomes. Second, the prompt aims to ensure the model prioritizes certain objectives using capitalization. Third, the prompt uses a few examples, which is known to increase model performance.

**Analysis.** Our analysis proceeds in three phases. First, we test the ability of smart readers to simplify the contracts and policies using objective metrics. As detailed in part IV.A, these included length, complexity, and readability. Second, we assess the quality of the summaries and the extent to which they capture key information. To that end and as we explain in more detail below,<sup>42</sup> we use Spotify’s Terms of Use as a sample contract and identified specific important

---

<sup>40</sup> Chunking is not a trivial task, since cutting off a document in the middle risks disrupting its meaning. This risk is especially true for cutting off in the middle of a sentence, but is also true for other cutting criteria given textual inter-dependence. Our chunking algorithm divided the document into sentences and then made sure each chunk only had full sentences. A more robust system would have divided the document into clauses, but besides the technical difficulty of detecting clause limits, even this approach would cut off inter-clause dependencies.

<sup>41</sup> *Introduction | LangChain*, LANGCHAIN, <https://python.langchain.com/en/latest/index.html> (last visited July 24, 2023) [<https://perma.cc/LM58-BUSQ>].

<sup>42</sup> See Part IV.B.

contractual aspects and possible consumer traps. We then assess the extent to which two simplification outputs included the issues we spotted in their summary. Third, we examine in-depth the simplification of eight specific clauses. Whereas Part V focuses on the high-level results, Part V outlines the analysis process of the specific terms, its results, and our evaluation.

#### IV. HIGH-LEVEL RESULTS

We now detail our general results. In Section A, we discuss the simplification assessments. Using common tools, we objectively measure text length, complexity, and readability. We supplement these objective metrics in Section B, where we detail our subjective impression of two of the simplifications and their ability to capture key information.

##### *A. Simplification Assessment*

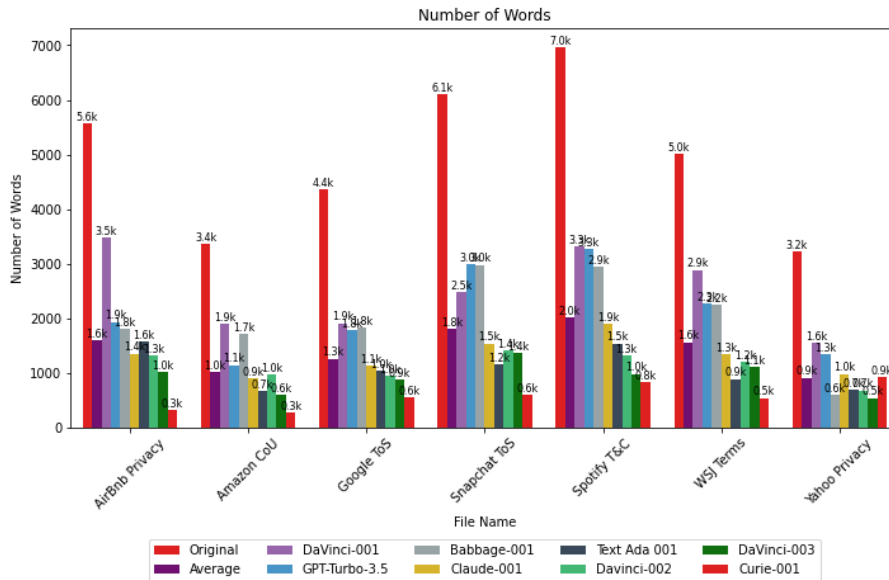
###### 1. Text Length

We started by measuring the reduction in words. On average and across all contracts, the various models produced a text that was about 30% of the length of the original in terms of words. In terms of reading time,<sup>43</sup> if the original version would take on average 20 minutes and 45 seconds to read, the simplified version only takes 6 minutes and six seconds, a time-saving of 14 minutes and 39 seconds. The following figure summarizes the average effect across all agreements and models.

---

<sup>43</sup> Based on Marc Brysbaert, *How Many Words Do We Read Per Minute? A Review and Meta-Analysis of Reading Rate*, 109 J. MEMORY & LANGUAGE 1, 21 (2019) (the average adult reads at a rate of 238 words per minute for non-fiction texts and 260 words per minute for fiction ones). As suggested to us by Professor Tim Samples, the reading time for difficult texts is longer, and by the same source it is estimated as 238 multiplied by 4.6 divided by the mean word length. The following Figure includes this method of analysis.

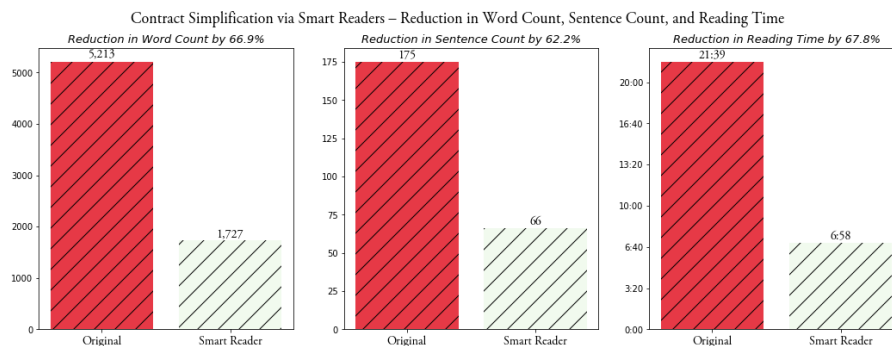
**Figure 1: Length Reduction (in Thousand Words)**



Another notable feature of the models is their great variability. Despite employing a similar prompt, the models produced wildly different results. While all models did well in terms of summarization, the worst one (Davinci-001) saw a reduction of 49%, while the best one (Curie-001) decreased 88.4% of the original length. And while all models perform admirably, the longest version was three times longer than the shortest one. In other words, there is a large degree of inconsistency between models.

The following figure aggregates the reduction across all documents to provide an overview of the average reduction in the number of words and sentences, as well as the required reading time.

**Figure 2: Aggregated Reduction Results: Words, Sentences & Reading Time**



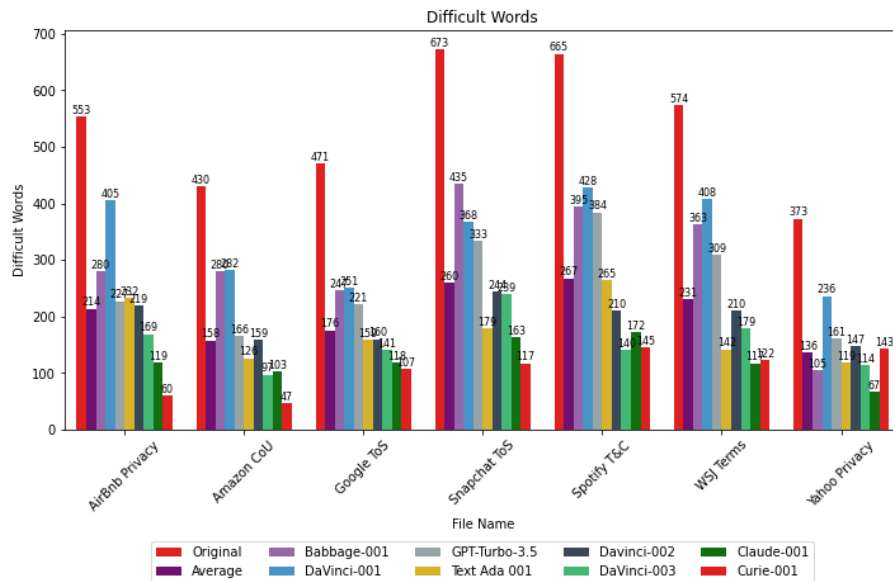
2. Text Complexity

One way to assess the simplification of text is by a count of difficult words in the text before and after simplification. Although there is no single way to

measure word difficulty, we relied on the word list compiled for a famous readability test: Dalle-Chall.<sup>44</sup> This list is admittedly somewhat limited and not tailored to legal jargon. However, it is adequate for our purposes of assessing differences between versions (rather than the absolute number of difficult words).

The Figure below shows the number of difficult words in the different documents. The red bar shows the number of difficult words in the original text and the purple bar summarizes the average number of difficult words across the various models. Across all documents, we note an average reduction of 328 difficult words, representing a 61% reduction. As before, there is great model variability in simplification. Still, even the worst model (DaVinci-001) substantially reduced the number of difficult words, removing 36% of them.

**Figure 3: Difficult Words**



### 3. Text Readability

There is more to text readability than difficult words, and the literature on readability has developed several quantitative tools to measure readability.<sup>45</sup> As discussed above, we report on the results of the famous Flesch-Kincaid score and the average of a number of readability measures (CRM).<sup>46</sup>

<sup>44</sup> See Common Education Data Standards, *supra* note 37. We implemented this word list via the Textstat library in Python.

<sup>45</sup> See *id.* (listing out different readability tests).

<sup>46</sup> The results stem from the mean of popular Python libraries that implement readability tests: textacy, textstat, textexplore, readcalc, pylexitext, and readability. Chaitanya Aggarwal & Shivam Bansal, *Textstat 0.7.3*, PYTHON PACKAGE INDEX (Mar. 15, 2022), <https://pypi.org/project/textstat/> [<https://perma.cc/2LUW-PJ29>]; Temiloluwa Awoyele, *Text-Explore 0.0.2*, PYTHON PACKAGE INDEX (Mar. 18, 2022), <https://pypi.org/project/text-explore/> [<https://perma.cc/RGA9-Q8ME>]; Victor Bona, *Pylexitext 0.3.1*,

As the next figure depicts, the original contracts were written at a level that requires between 10 and 14 years of schooling on the Flesch-Kincaid measure.<sup>47</sup>

#### Figure 4: Text Readability Flesch-Kincaid

On average, LLMs reduced the reading difficulty of the original contracts by 1.47 grade levels. That said, there was great variability among the models, with the best performing model (Claude-001) reducing the reading level by an average of 5.6 grade levels, down to a 5.4 grade level. This reduction would make contracts accessible to 11-year-olds. This improvement is quite important as a large body of scholarship recommends that reading materials be accessible to people who read between the sixth to eighth grade level.<sup>48</sup>

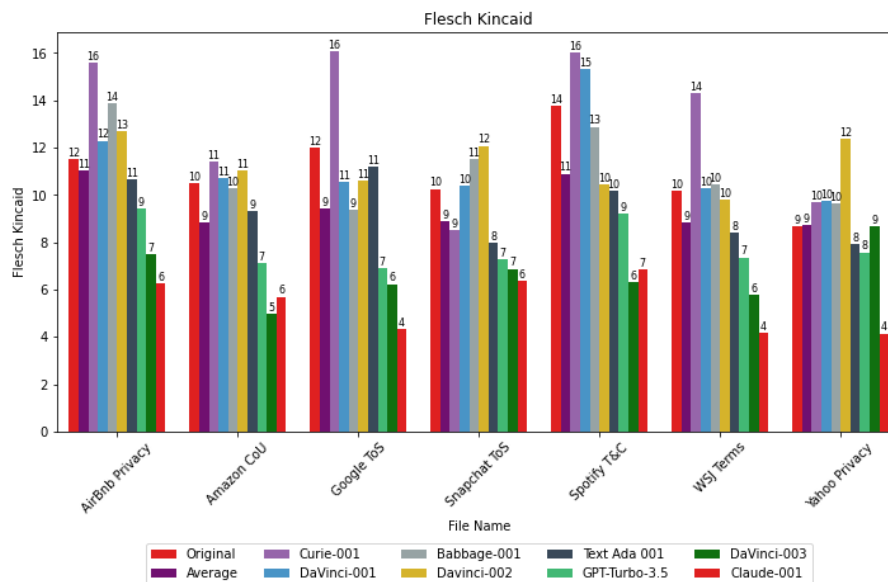
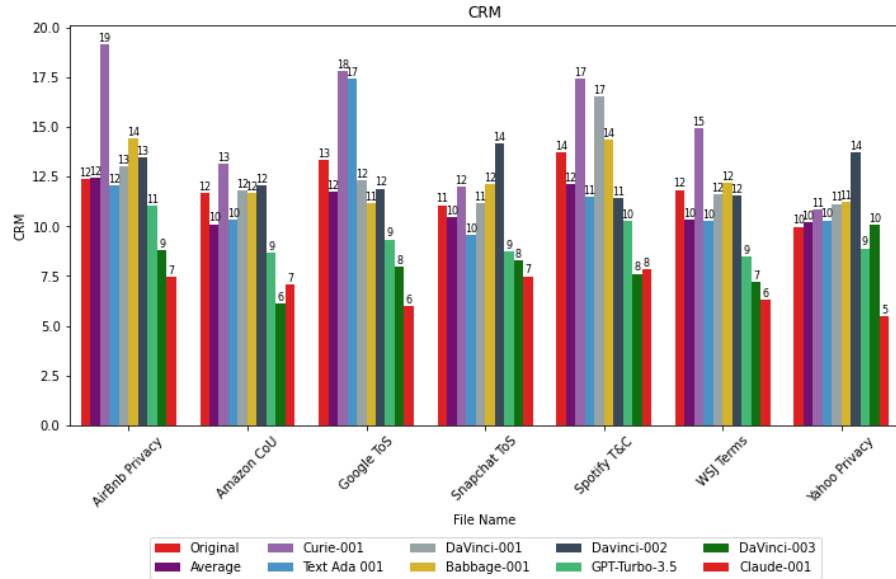


Figure 5 below provides similar numbers with the CRM measure. The average reduction on the CRM was more modest, with close to a single grade level. However, here too, the best performing model (Claude-001) did very well and reduced the reading level by 5 grade years, down to a level close to the seventh grade (that is, a 6.8 grade level).

PYTHON PACKAGE INDEX (May 19, 2021), <https://pypi.org/project/pylexitext/> [https://perma.cc/BP4Y-WPV7]; Burton DeWilde, *Textacy 0.13.0*, PYTHON PACKAGE INDEX (Apr. 2, 2023), <https://pypi.org/project/textacy/> [https://perma.cc/2359-TRPJ]; Joao Palotti, *ReadabilityCalculator 0.2.37*, PYTHON PACKAGE INDEX (Apr. 30, 2018), <https://pypi.org/project/ReadabilityCalculator/> [https://perma.cc/58RE-ZX24]; Andreas van Cranenburgh, *Readability 0.3.1*, PYTHON PACKAGE INDEX (Jan. 12, 2019), <https://pypi.org/project/readability/> [https://perma.cc/D3PD-UNQ9].

<sup>47</sup> For reasons of legibility, we rounded the labels on the bars, though their height reflects their unrounded score.

<sup>48</sup> See, e.g., Kristie B. Hadden, Latrina Prince, Laura James, Jennifer Holland, & Christopher R. Trudeau, *Readability of Human Subjects Training Materials for Research*, 13 J. EMPIRICAL RSCH. ON HUM. RSCH. ETHICS 95, 96 (2018) (noting that “experts recommend that written materials developed for public use are written at a sixth to eighth grade level or below for ease of reading and comprehension”).

**Figure 5: Text Readability CRM**

Quite surprisingly, despite our prompts, some models have made them *more* complex. Yet again, the models have had mixed success in simplifying the contracts. The most consistent models were Claude-001 and Text-Davinci-003, a finding consistent with them being the most advanced in our group.

### B. Quality Assessment

To develop a better sense of the *quality* of the outputs generated by LLMs, we supplemented our metric-based examination above with a more subjective evaluation. Essentially, we compared the first parts<sup>49</sup> of an original text—Spotify’s Terms of Use—with two simplified outputs, produced by ChatGPT-Turbo and Claude. Particularly, we sought to examine whether the simplified versions captured the key issues and points we identified in the original text.

We started by reading the original text and highlighting clauses likely to be of special importance for unwary consumers. We detected 11 central or tricky points (dubbed here “traps”). We then read the two simplified versions, examining whether these versions properly mentioned the identified 11 key points. We also noted our general subjective impression of the texts’ quality, presentation, and user-friendliness.

<sup>49</sup> By first parts, we mean clauses one to three, which contained 2,360 words.

Overall, we found that both platforms made the contracts much simpler and captured most of the important information. However, significant differences were evident between the two outputs in terms of length, visual presentation, and the use of bullet points. The visual presentation, however, should not be emphasized. It was more an artifact of the way we coded the smart reader than a technological challenge. With more resources, we can produce a smart reader that will organize the text, use bullet points, and even engage in creative design, including colors and comic-style graphics.

In addressing the 11 traps, both models performed reasonably well, albeit with somewhat varying degrees of success. Both models captured most, but not all, traps and important information in their summaries. For example, ChatGPT Turbo included 9 of the 11 traps in its summary. Interestingly, one of the traps that Claude omitted overlapped with one of these two missing traps. This finding has two implications. First, if some models fail to address certain issues, another model better at detecting them can supplement them (the ‘ensemble’ method). Second, there is room for research into the possibility that current LLMs might systematically miss some types of information.

In general, we observe that some of the omissions and presentation issues we encountered are due to the need to cut the original texts into smaller chunks. This “chunking” strategy is common today, but it is imperfect.<sup>50</sup> It interferes with the flow of the text and cuts context, an especially important concern in interdependent contracts. With time, input length will not pose as much of an issue, so the flaws associated with this specific problem are transient.

This broad review allowed us to develop a general sense of quality. However, it did not capture some of the subtleties of simplification. To that end, we next analyzed specific clauses.

## V. SIMPLIFICATION & QUALITY: SPECIFIC CLAUSES

To develop a more robust understanding of the capability of smart readers to produce high-quality simplifications, we next shifted our focus from the entire contracts to specific clauses within them. We chose to focus on eight specific terms that may pose challenges for the unsuspecting consumer. That is, rather than measuring performance on random clauses, some of them potentially trivial,

---

<sup>50</sup> Alberto Artasánchez & Prateek Joshi, *Assessing Text Through Bag of Words Model in Natural Language Processing*, MEDIUM (June 9, 2020), <https://odsc.medium.com/assessing-text-through-bag-of-words-model-in-natural-language-processing-9654f17abf56> [<https://perma.cc/VY45-KSG2>].

we actively selected clauses that we thought mattered for consumers. These included, for example, clauses that address cancelation, information sharing, liability limitation, dispute resolution, and unilateral modifications.

Because this focus resulted in relatively short texts (i.e., specific clauses), we could effectively use GPT-4, the current state-of-the-art model. For consistency, we used the same prompt as before. In what follows, we delineate the contract clauses we examined. For each clause, we provide the original text, followed by the simplified version offered by GPT-4. We then evaluate the simplification in terms of length, complexity/readability, and quality.

#### A. *Wall Street Journal: Changes to Subscriber Agreement*

##### Original<sup>51</sup>

**1. Changes to Subscriber Agreement:** We may change the terms of this Agreement at any time by notifying you of the change in writing or electronically (including without limitation, by email or by posting a notice on the Service that the terms have been “updated” or similar words). The changes also will appear in this document, which you can access at any time by going to the Subscriber Agreement and Terms of Use link at the footer of those Services which are made available through a website or to the Legal or Legal Notices area of those Services which are made available as an application. By using a Service after changes are made to this Agreement you signify that you agree to be bound by such changes.

##### Simplified

**1. Subscriber Agreement Updates:** We can update this Agreement whenever needed. We’ll let you know about these updates, either in writing or digitally (like by email or posting an update notice on the Service). These updates will be added to this document. You can find it anytime by clicking on the Subscriber Agreement and Terms of Use link at the bottom of any website-based Services, or in the Legal area of any app-based Services. If you keep using a Service after we update this Agreement, it means you accept these updates.

##### Evaluation

This clause allows the service provider to change the terms of the agreement unilaterally, an issue of considerable import for consumers.<sup>52</sup> We find the simplification to be highly effective. It uses the second person to refer to the

<sup>51</sup> *WSJ Terms*, *supra* note 33.

<sup>52</sup> *E.g.*, Shmuel I. Becher & Uri Benoliel, *Sneak In Contracts*, 55 GA. L. REV. 657, 657 (2021) (asserting that unilateral modifications can “affect virtually every aspect of a contract”); *see also* David Horton, *The Shadow Terms: Contract Procedure and Unilateral Amendments*, 57 UCLA L. REV. 605, 645 (2010) (describing how unilateral amendments give drafters “a kind of contractual dexterity”).

reader. It starts the paragraph with a thesis statement (as in TRIAC and other writing conventions)<sup>53</sup> and offers a direct summary of the effect of the clause.

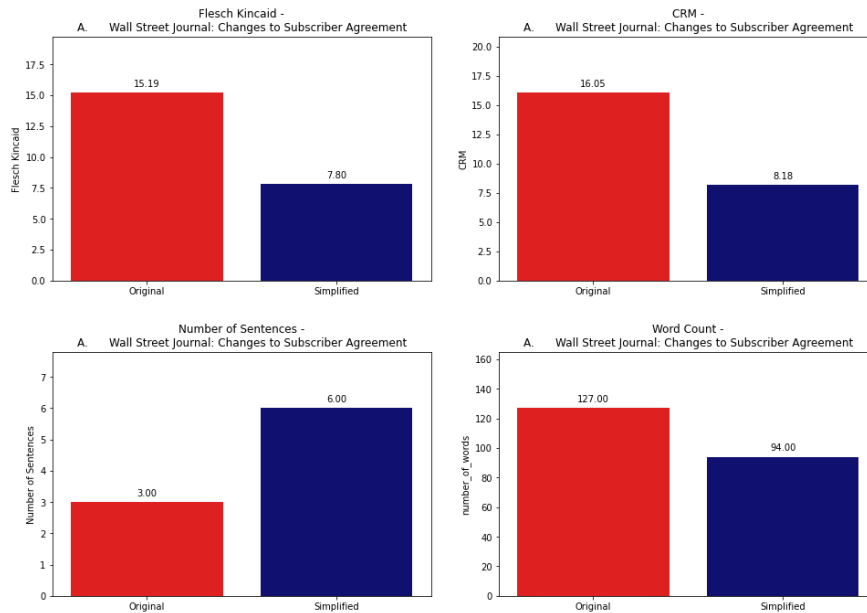
While generally true to the original, there is yet a subtle issue here. In the original, changes to the contract are only effective if communicated. In the simplified version, it is implied that changes are done unilaterally, and some users may infer that communicating those changes is more of a courtesy than a prerequisite. If a firm makes uncommunicated changes to its agreement, the simplified version may mislead the consumer into thinking that the changes are effective. Still, the difference is not large, and the actual meaning of “we will let you know about these updates” (as per the simplified version) may be somewhat ambiguous.

Either way, the rest of the paragraph is well-executed. The quantitative analysis depicted in Figure 6 shows a reduction of nearly 7.5 grade levels on the Flesch-Kincaid measure and 8 grade reduction on the average of scores of the various readability measures. In either case, the text is evaluated as readable by an eighth-grader. Furthermore, the simplified version reduced 26% of the text, cutting the number of words by 33 (out of 127).<sup>54</sup> The number of sentences doubled from three to six due to splitting long sentences into shorter ones. Consistently, the average word length, although not depicted (in this and the following figures), falls from 5.17 characters to 4.45.

---

<sup>53</sup> Eric Drown, *TRIAC Paragraph Structure*, UNIV. NEW ENG., <https://ericdrown.uneportfolio.org/triac/> (last visited July 24, 2023) [<https://perma.cc/SD4R-PQKQ>].

<sup>54</sup> Counting words depends on a technique involving the splitting of words called tokenization. John Maeada & Matthew Bolanos, *What Are Tokens?*, MICROSOFT (May 23, 2023), <https://learn.microsoft.com/en-us/semantic-kernel/prompt-engineering/tokens> [<https://perma.cc/S8XQ-5RCD>]. Microsoft Word counts words in a fairly simplistic manner, counting contractions and hyphenated words as a single word, thus somewhat biasing results. Carol Bratt, *Ignore Hyphens When Performing a Word Count in MS Word*, DAVE'S COMPUT. TIPS (Oct. 23, 2012), <https://davescomputertips.com/ignore-hyphens-when-performing-a-word-count-in-ms-word/> [<https://perma.cc/B4MK-K6QY>].

**Figure 6: Clause Simplification, WSJ***B. Wall Street Journal: Agreement to Arbitrate***Original<sup>55</sup>****14. Agreement to Arbitrate.**

14.1 The parties acknowledge that any statutory or common law claims related to intellectual property may require forms of equitable relief that are best administered by courts; accordingly, the parties agree that except for statutory or common law claims related to intellectual property and disputes that qualify for small claims court, any controversy or claim arising out of or relating to this Agreement or any aspect of the relationship between us, whether based in contract, tort, statute, fraud, misrepresentation or any other legal theory, will be resolved by arbitration administered by the American Arbitration Association (“AAA”) in accordance with its Commercial Arbitration Rules and the Supplementary Procedures for Consumer Related Disputes. Judgment on the award rendered by the arbitrator may be entered in any federal or state court of competent jurisdiction located in the County of New York in the State of New York. For more information about arbitration, the AAA and the arbitration process, please consult the American Arbitration Association web site at [adr.org](http://adr.org). You agree that by entering into this Agreement, you and we are each waiving the right to trial by jury, except as otherwise stated above. Any arbitration under this Agreement will take place on an individual basis; class arbitrations and class actions are not permitted and, by entering into this Agreement, you are giving up the ability to participate in a class action. All issues are for the arbitrator to decide, except that issues relating to the scope, application, and enforceability of

<sup>55</sup> *WSJ Terms*, *supra* note 33.

the arbitration provision are for the court to decide. As stated in Section 15 below, New York law applies to any arbitration under this section, but the parties acknowledge that the Federal Arbitration Act governs the interpretation and enforcement of this provision. This agreement to arbitrate shall survive termination of this Agreement. This arbitration agreement does not preclude you from bringing issues to the attention of federal, state, or local agencies, and, if the law allows, they can seek relief against us for you.

14.2 Unless you and we agree otherwise, the arbitration will take place either in the subscriber's county of principal residence or the County of New York in the State of New York. For claims of \$14,000 or less, you can choose whether you would like arbitration carried out based only on documents submitted to the arbitrator, or by a hearing in person, or by phone.

14.3 The arbitrator may award money or equitable relief in favor of only the individual party seeking relief and only to the extent necessary to provide relief warranted by that party's individual claim. To reduce the time and expense of the arbitration, the arbitrator will not provide a statement of reasons for his or her award unless requested to do so by both parties. Unless both you and we agree otherwise, the arbitrator may not consolidate more than one person's claims, and may not otherwise preside over any form of a representative or class proceeding.

### Simplified

#### 14. Solving Disputes.

14.1 We agree that some issues, like those tied to ideas, may need to be sorted out by courts. Except for these, and for small disputes that fit into small claims court, we'll deal with any problems or disputes from this Agreement through a process called arbitration. This will be managed by the American Arbitration Association ("AAA"). You can find more about this at [adr.org](http://adr.org). Decisions from arbitration can be used in any federal or state court in New York County, State of New York. By agreeing to this, both you and we give up the right to a jury trial, except for the cases we just noted. The arbitrator will deal with issues one by one; group arbitrations and group actions aren't allowed. We agree that this arbitration process, based on New York law, will still apply even after this Agreement ends.

14.2 Unless we both say otherwise, the arbitration will happen either where you live or in New York County, State of New York. For claims of \$14,000 or less, you get to pick how the arbitration will be done: only by looking at submitted documents, through a meeting in person, or by phone.

14.3 The arbitrator can award money or make fair changes in favor of only the person asking for it and only to the level needed to provide help based on their claim. The arbitrator won't provide a reason for the decision unless both parties ask. The arbitrator won't deal with more than one person's claims at a time unless both you and we agree otherwise.

## Evaluation

This clause provides for mandatory arbitration and makes specific procedural requirements on a highly sensitive issue that impacts access to justice, corporate power, and the development of the law.<sup>56</sup> While the simplification is generally good, we note a few problems. First, perhaps by virtue of the prompt's insistence on a change of terms, it replaced the heading 'Agreement to Arbitrate' with 'Solving Disputes.' While the latter communicates better, it does not make the term arbitration sufficiently salient.

Relatedly, some choices of words may implicate different legal consequences. The model translated intellectual property to "ideas." This change is quite wrong: not all ideas are subject to IP, and this may give a sense that more disputes belong in courts than the original agreement allows. Yet again, the problem here may stem from our prompt, which insisted on using simpler terms. A similar issue afflicts the choice of "group actions" instead of "class actions" and the language suggesting that arbitration awards can be "used" rather than "entered."

In terms of quality, a worrisome omission is that the original clarifies that the customer may still bring complaints to state or federal agencies. The simplified version does not mention that. This omission can deprive the customer of important rights.

A final issue is that the model translated equitable relief to "fair changes." These two ideas differ, and the simplified version could mislead. Yet, the dilemma of simplification of this term is quite pressing, and the average consumer may find both terms ambiguous. It is hard to effectively translate equitable relief—which may consist of in-kind remedies, injunctive remedies, apologies, and other measures—to a simple term. At the same time, keeping the original term would sacrifice the ability of laypeople to parse it. All in all, this simplification is disputable but not irrational.

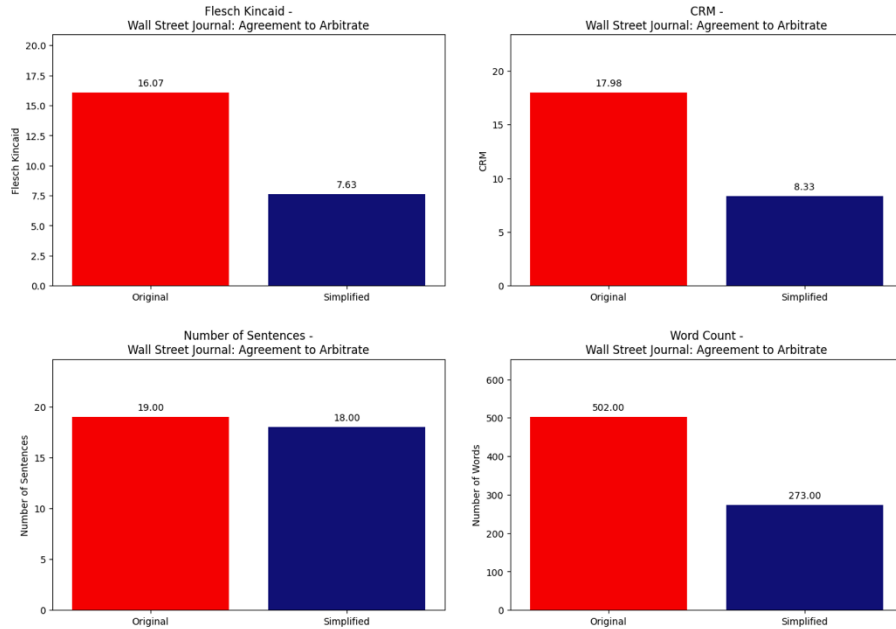
On the quantitative measures, we see a dramatic reduction in grade level evaluated for the reading of the agreement. Again, the simplification transformed an agreement readable by PhDs into one readable by eighth-graders. The word

---

<sup>56</sup> David Horton & Andrea Cann Chandrasekher, *After the Revolution: An Empirical Study of Consumer Arbitration*, 104 GEO. L.J. 57, 57 (2015) (describing how the stakes around this issue have "soared" since 2010); Jeff Sovern, Elayne E. Greenberg, Paul F. Kirgis, & Yuxiang Liu, "Whimsy Little Contracts" with Unexpected Consequences: An Empirical Analysis of Consumer Understanding of Arbitration Agreements, 75 MD. L. REV. 1, 2-3 (2015) (finding only a quarter of more than 5,000 respondents had a correct understanding of arbitration agreements); see also Shmuel I. Becher & Uri Benoliel, *Dark Contracts*, 64 B.C. L. REV. 55, 68-71 (2023).

count has fallen by 46%. The number of sentences has only fallen slightly (from 19 to 18), and the average word length fell from 5.04 to 4.33.

**Figure 7: Clause Simplification, WSJ (2)**



### C. Airbnb: Collecting Personal Information from Third Parties

#### Original<sup>57</sup>

##### 2.4 Personal Information We Collect from Third Parties

We collect personal information from other sources, such as:

- **Third-Party Services.** If you link, connect, or login to the Airbnb Platform with a third-party service (e.g., Google, Facebook, WeChat), you direct the service to send us information such as your registration, friends list, and profile information as controlled by that service or as authorized by you via your privacy settings at that service.
- **Background Information.** For Members in the United States, to the extent permitted by applicable laws, we may obtain, for example, reports of criminal records, sex offender registrations, and other information about you and/or your background. For Hosts in India, to the extent permitted by applicable laws, we may perform criminal background checks. For Members outside of the United States, to the extent permitted by applicable laws and with your consent where required, we may obtain police, background, or registered sex offender checks. We may use your information, including your full name and date of birth, to obtain such reports. [...]

<sup>57</sup> Airbnb Privacy Policy, *supra* note 33.

- Referrals and Co-Travelers. If you are invited to the Airbnb Platform, for example, as a co-traveler on a trip, the person who invited you can submit personal information about you such as your email address or other contact information.
- Guest Travel Insurance. If you make a claim under our Guest Travel Insurance policy, we will receive information regarding your claim in order to process, handle, or assess your claim and as outlined in this Privacy Policy.
- Other Sources. To the extent permitted by applicable law, we may receive additional information about you, such as references, demographic data, or information to help detect fraud and safety issues from third-party service providers and/or partners, and combine it with information we have about you. For example, we may receive background check results or fraud warnings from identity verification service providers for use in our fraud prevention, security investigation, and risk assessment efforts. We may receive information about you and your activities on and off the Airbnb Platform, or about your experiences and interactions from our partners. We may receive health information, including but not limited to, health information related to contagious diseases.

### **Simplification**

2.4 Info We Get from Other Sources. We get personal info from other places like:

- Other Services. If you connect your Airbnb account with a service like Google, Facebook, or WeChat, you allow them to send us certain info like your registration and profile details.
- Background Checks. In the US, we might get reports on criminal records and other background info. In India, we might run criminal background checks on hosts. In other countries, we might get police or background checks if the law allows and you agree.
- Invites and Co-travelers. If someone invites you to Airbnb, like to join a trip, they might give us your contact info.
- Guest Travel Insurance. If you make a claim under our travel insurance policy, we get info about your claim.
- Other Places. If the law allows, we might get more info about you from other service providers or partners, and mix it with the info we already have. This could be background checks, fraud warnings, info about your activities on and off Airbnb, and even health info related to infectious diseases.

### **Evaluation**

This provision delineates the categories of personal data that Airbnb collects from third-party sources, which is another important practice for consumers to understand. The simplification has successfully distilled the original provision's complex legal language into more accessible terms without distorting the overall

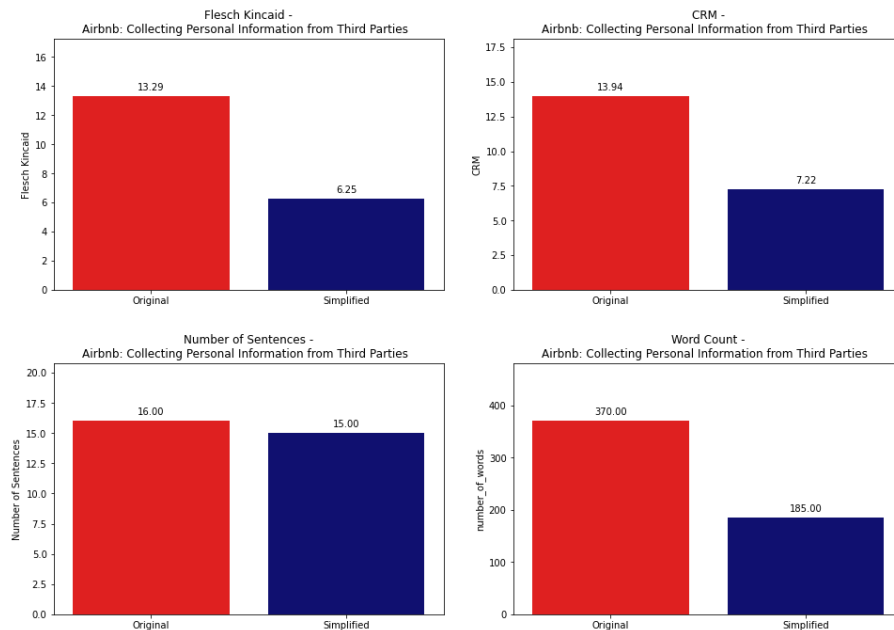
meaning. The use of “other sources” instead of third-party services is quite helpful, and the use of “info” instead of information may be grammatically inaccurate yet useful (or at least understandable).

However, certain points could be clearer and more comprehensive. For instance, the simplified version omits the example that some of the information shared may contain the user’s friends list. This omission is not misleading per se, but we suspect that consumers may not otherwise expect Airbnb to share this information with third-parties.

As for background checks, the simplification converts “with your consent where required” to a proposition that implies that consent will always be sought. This modification is misleading because sharing will sometimes occur without consent. An arguably more minor omission regards information sharing with the insurer. The simplified version does not mention that the privacy policy limits information sharing; i.e., it gives a sense that information sharing is unlimited.

The quantitative measure shows another substantial improvement, reducing the readability level from college level to sixth or seventh grade. Furthermore, the word count has declined by 50%: from 370 words to 185. This drop included a minimal decrease in the number of sentences (from 16 to 15) and moderate reduction in the average word length (from 5.04 to 4.33).

**Figure 8: Clause Simplification, Airbnb**



#### D. Netflix: Cancellation

##### Original<sup>58</sup>

3.4. Cancellation. You can cancel your Netflix membership at any time, and you will continue to have access to the Netflix service through the end of your billing period. To the extent permitted by the applicable law, payments are non-refundable and we do not provide refunds or credits for any partial membership periods or unused Netflix content. To cancel, go to the “Account” page and follow the instructions for cancellation. If you cancel your membership, your account will automatically close at the end of your current billing period. To see when your account will close, click “Billing details” on the “Account” page. If you signed up for Netflix using your account with a third party as a Payment Method and wish to cancel your Netflix membership, you may need to do so through such third party, for example by visiting your account with the applicable third party and turning off auto-renew, or unsubscribing from the Netflix service through that third party.

##### Simplification

3.4. Ending Your Membership. You can stop your Netflix membership whenever you want. You’ll still be able to use Netflix until the end of your billing cycle. Under current law, we can’t refund payments and we don’t give refunds or credits for part of a membership period or for unused Netflix content. To stop your membership, visit the “Account” page and follow the steps there. If you stop your membership, your account will close at the end of your billing cycle. To find out when that is, click “Billing details” on the “Account” page. If you joined Netflix using a third party for payment and you want to stop your membership, you may have to do this through that third party. This might involve turning off auto-renew or unsubscribing from Netflix through them.

##### Evaluation

This provision governs the cancellation of a Netflix subscription, yet another tricky issue in consumer contracts.<sup>59</sup> The simplification seems to maintain the overall integrity of the original paragraph while making it more accessible to the reader. However, there is a misinterpretation of the refund policy that could potentially mislead consumers.

The original text states that Netflix’s general policy is not to provide refunds or credits for any partial membership periods or unused Netflix content, but that

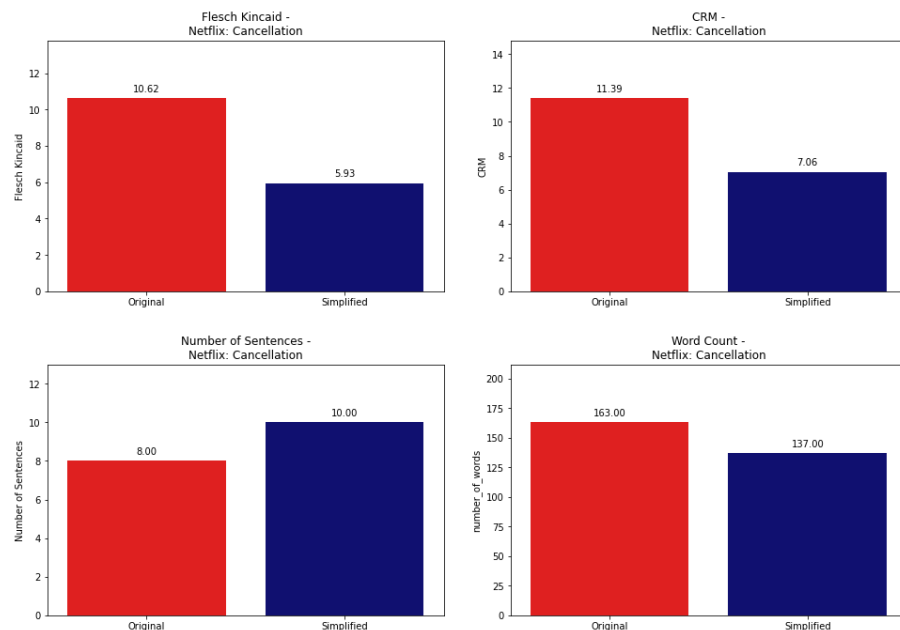
<sup>58</sup> *Netflix Terms*, *supra* note 33.

<sup>59</sup> Cf. Roman Inderst & Marco Ottaviani. *Sales Talk, Cancellation Terms and the Role of Consumer Protection*, 80 REV. ECON. STUD. 1002, 1002 (2013) (discussing similar concerns in insurance plans and annuities).

the firm can make exceptions to the extent permitted by applicable law. This text suggests that in some jurisdictions, Netflix may indeed be *required to issue refunds*. In contrast, the simplified text asserts that “under current law, we can’t refund payments.” This language indicates that Netflix is legally *prohibited from providing refunds*. This erroneous interpretation potentially miscommunicates Netflix’s refund policy to consumers, particularly those residing in regions where laws may, in fact, mandate refunds. This difference is a major issue that can mislead consumers and undermine their rights.

The quantitative measure shows another significant improvement, setting the readability level at between the sixth and seventh grade (instead of the high school level required for reading the original). At the same time, the word count has only declined by 16% (from 163 to 137). This included a small increase in the number of sentences (from 8 to 10), and a slight decrease in the length of words (from 4.82 to 4.47).

**Figure 9: Clause Simplification, Netflix**



*E. Amazon: Reviews, Comments, Communications, & Other Content*

**Original<sup>60</sup>**

REVIEWS, COMMENTS, COMMUNICATIONS, AND OTHER CONTENT You may post reviews, comments, photos, videos, and other

<sup>60</sup> *Amazon Terms*, *supra* note 33.

content; send e-cards and other communications; and submit suggestions, ideas, comments, questions, or other information, so long as the content is not illegal, obscene, threatening, defamatory, invasive of privacy, infringing of intellectual property rights (including publicity rights), or otherwise injurious to third parties or objectionable, and does not consist of or contain software viruses, political campaigning, commercial solicitation, chain letters, mass mailings, or any form of “spam” or unsolicited commercial electronic messages. You may not use a false email address, impersonate any person or entity, or otherwise mislead as to the origin of a card or other content. Amazon reserves the right (but not the obligation) to remove or edit such content, but does not regularly review posted content.

If you do post content or submit material, and unless we indicate otherwise, you grant Amazon a nonexclusive, royalty-free, perpetual, irrevocable, and fully sublicensable right to use, reproduce, modify, adapt, publish, perform, translate, create derivative works from, distribute, and display such content throughout the world in any media. You grant Amazon and sublicensees the right to use the name that you submit in connection with such content, if they choose. You represent and warrant that you own or otherwise control all of the rights to the content that you post; that the content is accurate; that use of the content you supply does not violate this policy and will not cause injury to any person or entity; and that you will indemnify Amazon for all claims resulting from content you supply. Amazon has the right but not the obligation to monitor and edit or remove any activity or content. Amazon takes no responsibility and assumes no liability for any content posted by you or any third party.

### **Simplification**

**POSTS, MESSAGES, AND OTHER CONTENT** You can post reviews, comments, photos, videos, and more. You can send messages and share ideas, comments, questions, or other info. But make sure your content is legal, polite, respects others’ privacy, and doesn’t infringe on anyone’s rights. Don’t post harmful or unwanted content like spam, viruses, false info, or anything that misleads others. We can remove or change this type of content, but we don’t check all content regularly.

If you do post or share stuff, unless we say otherwise, you’re giving Amazon permission to use it. This permission doesn’t end, doesn’t cost anything, and can be passed on. We can use it, change it, publish it, perform it, translate it, make new stuff from it, and share it anywhere in any form. You also let Amazon and others we give permission to use your name with your content, if they want. You promise that you own or control the rights to what you post, that it’s correct, that it won’t break this rule or hurt anyone or anything, and that you’ll cover Amazon for all claims related to your content. Amazon can monitor, change, or remove any activity or content but isn’t required to. Amazon isn’t responsible for any content posted by you or anyone else.

## Evaluation

This provision governs content contribution to the platform, most notably in the form of reviews. This provision raises an important issue since consumers often rely on online reviews and use them in their decision-making. Hence, ensuring reliable and robust online information flows is crucial in disciplining sellers and maintaining efficient markets.<sup>61</sup>

Overall, we find the simplification above to be effective. However, some issues require attention. First, the simplified version leaves out a few important restrictions. It fails to mention the prohibition on email spoofing and political campaigning explicitly mentioned in the original, thereby providing less comprehensive guidance to users. At the same time, the original possibly contained too many illustrations, so the balance is delicate.

Second, the contract contains an important clause about the warranties provided by users when they post content, which is especially weighty in its implications. Users have to guarantee the accuracy of their reviews and ensure that these reviews will not harm others. This requirement could be burdensome to users and is quite unexpected, given that truthful reviews may well harm unscrupulous, negligent, or underperforming sellers. The simplification communicates most of it, but perhaps a more formal tone might be beneficial in highlighting the gravity of these obligations. This difference potentially points to a deeper problem at the heart of the simplification project: whereas simplification often entails a more casual or flippant tone, formality is sometimes a good signal of the gravity of obligations.<sup>62</sup>

By readability metrics, this was the most complex original text so far. The reduction from 19-20 years of education to a level readable by an eighth grader is, therefore, quite impressive. Also, the length fell by 26% (from 304 words to 226). This dip was nonetheless accompanied by an increase in the number of sentences (from 8 to 12), which, as noted, often results from text simplification

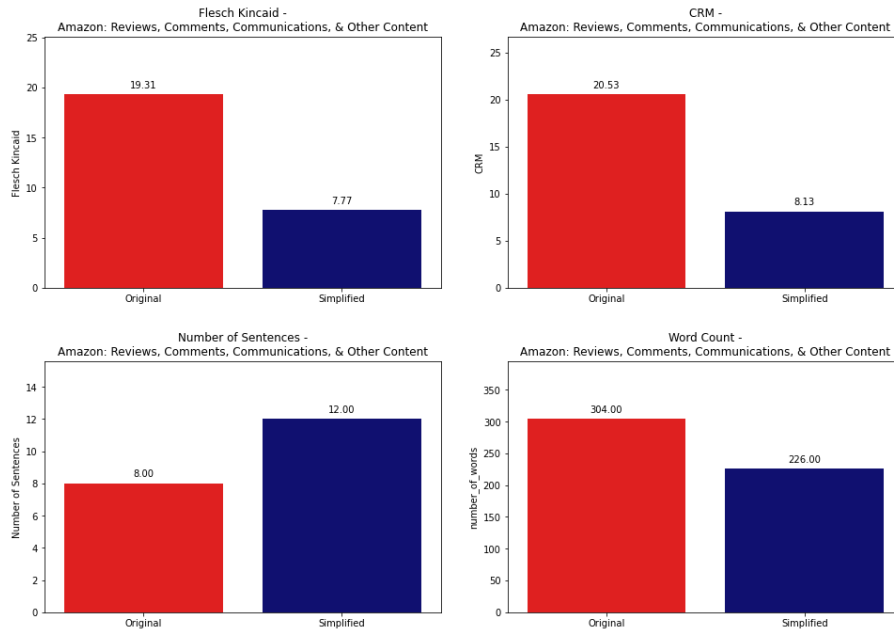
---

<sup>61</sup> See Shmuel I. Becher & Tal Z. Zarsky, *E-Contract Doctrine 2.0: Standard Form Contracting in the Age of Online User Participation*, 14 MICH. TELECOMMS. & TECH. L. REV. 303, 333 (2008) (viewing accreditation issues in terms of reliability and trustworthiness as challenges to online information flow); Arbel & Shapira, *supra* note 8, at 952 (asserting “reputational information is too often unreliable”); Yonathan A. Arbel, *Reputation Failure: The Limits of Market Discipline in Consumer Markets*, 54 WAKE FOREST L. REV. 1239, 1275 (2019) (explaining why statistically reviews are unreliable as “sole guides for consumer decision-making”).

<sup>62</sup> Arbel and Toler test this and do not find evidence to support the red-flag theory of formality, yet in the context of a fairly specific intervention. See Arbel & Toler, *supra* note 9, at 896.

(i.e., employing short sentences). Finally, word length fell from 5.17 to 4.33 characters per word.

**Figure 10: Clause Simplification, Amazon**



#### F. Amazon: Risk of Loss

##### Original<sup>63</sup>

**RISK OF LOSS** All purchases of physical items from Amazon are made pursuant to a shipment contract. This means that the risk of loss and title for such items pass to you upon our delivery to the carrier. **RETURNS, REFUNDS AND TITLE** Amazon does not take title to returned items until the item arrives at our fulfillment center. At our discretion, a refund may be issued without requiring a return. In this situation, Amazon does not take title to the refunded item. For more information about our returns and refunds, please see our Returns Center.

##### Simplification

**LOSS RISK** When you buy physical items from Amazon, they are yours once we give them to the carrier for delivery.

**RETURNS, MONEY BACK, AND OWNERSHIP** If you return something, it's not ours until it gets to our fulfillment center. Sometimes, we might give you your money back without needing a return. If that happens, you still own the item. To learn more about returns and refunds, visit our Returns Center.

<sup>63</sup> *Amazon Terms*, *supra* note 33.

## Evaluation

This clause assigns the risk of delivery loss to the customer. Consumers may not expect such a clause, as they have little control over Amazon's shipping methods. Indeed, some jurisdictions regard contract terms that relieve the supplier of responsibility, which they would have to bear under law or by virtue of the contract (if not for that term), as presumably unfair.<sup>64</sup>

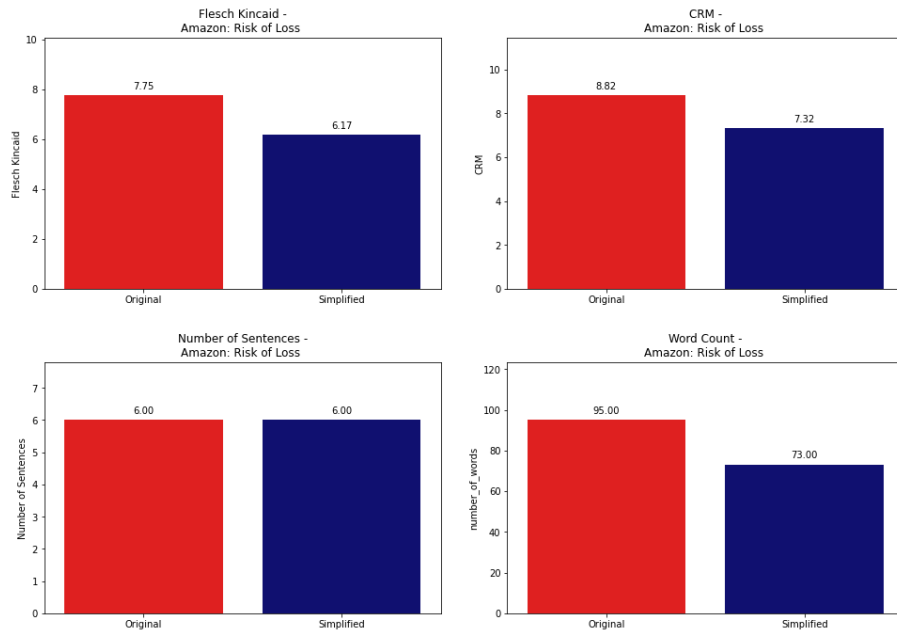
The simplification preserves the overall meaning, but it could do better in highlighting key customer responsibilities and exceptions. It effectively communicates that the items belong to the buyer upon delivery to the shipper. However, it does not communicate clearly enough that the customer bears the risk if something goes wrong with the delivery.

The same criticism, but to a lesser extent, applies to returns. The proposed simplification, "If you return something, it's not ours until it gets to our fulfillment center," does not adequately communicate the contractual risk allocation, such as what happens if items are lost in the return process. On the positive side, we note the effectiveness of using "you still own" relative to "take title."

Figure 11 depicts the most modest improvement in readability so far. Readability improved by just a single grade level, but that is likely because the original is both short and relatively simple. Still, we believe greater improvements could take place with additional prompting. Likewise, the change in length was also relatively less substantial, with a 23% reduction (from 95 words to 73). There was no change in the number of sentences and a marginal change to the word length, falling from 4.64 to 4.45.

---

<sup>64</sup> See, e.g., *Standard Contracts Law, 1982*, LEVITANSHARON & CO., <https://www.israelinsurancelaw.com/standard-contracts-law-1982/> (last visited July 24, 2023) [<https://perma.cc/6QNG-MJJX>] (listing out various sections of the contract code in Israel, particularly sections 3 and 4, which treat provisions that relieve supplier's responsibilities as per se unfair).

**Figure 11: Clause Simplification, Amazon (2)**

### G. Yahoo: Information Sharing

#### Original<sup>65</sup>

##### How We Share This Information

a. Yahoo shares information within its affiliated brands and companies. We also share information we have about you for the purposes described in this Privacy Policy, including to provide Services that you have requested (including when you connect with third-party apps and widgets). We do not sell, license or share information that individually identifies our customers with companies, organizations or individuals outside of Yahoo unless one of the following circumstances applies:

- i. With Your Consent. We will share information with companies, organizations or individuals outside of Yahoo when we have your consent.
- ii. Within Yahoo. Information may also be shared within Yahoo, including with other Yahoo Services and affiliates. Yahoo affiliates may use the information in a manner consistent with their privacy policies.
- iii. With Partners. We may share your information with nonaffiliated companies who are:
  - Trusted Partners. We provide user information to trusted partners who work on behalf of or with Yahoo based on our directions and in compliance with appropriate confidentiality measures. Learn more.

<sup>65</sup> Yahoo! Privacy Policy, *supra* note 33.

- Advertising, Analytics and Business Partners (Limited to Non-Personally Identifiable Information). We may share aggregated or pseudonymous information (including demographic information) with partners, such as publishers, advertisers, measurement analytics, apps, or other companies. For example, we may tell an advertiser how its ads performed or report how many people installed an app after seeing a promotion. We do not share personally identifiable information (like phone number or email address) with these partners, such as publishers, advertisers, ad agencies, or analytics partners.
- When you use third-party apps, websites or other products integrated with our Services, they may collect information about your activities subject to their own terms and privacy policies.
- We allow other companies that show advertisements on our webpages or apps to collect information from your browsers or devices. Other companies' use of cookies and other data collection technologies are subject to their own privacy policies, not this one. Like many companies, we may allow cookie matching with select partners. But, these parties are not authorized to access Yahoo cookies.

### Simplification

How We Share Your Info a. Yahoo shares your info within its related brands and companies. We also share info for the purposes mentioned in this Privacy Policy, including to give you Services you asked for. We don't sell or share info that identifies our customers with outside companies unless one of these situations happens:

- i. You Say Okay. We'll share info with outside companies if you give us permission.
- ii. Inside Yahoo. We might share your info within Yahoo, including with other Yahoo Services and linked companies. These Yahoo companies may use the info in line with their own privacy rules.
- iii. With Partners. We may share your info with companies who aren't part of Yahoo:
  - Trusted Partners. We give user info to trusted partners who do work for or with Yahoo. They follow our rules and keep the info confidential. Learn more.
  - Ad, Analytics, and Business Partners (Only Non-Personal Info). We may share grouped or pseudonymized info (like demographic info) with partners such as publishers, advertisers, and other companies. We don't share personal info (like phone number or email) with these partners.
  - When you use third-party apps or websites tied to our Services, they may collect info about what you do, based on their own rules and privacy policies.
  - We let other companies that show ads on our webpages or apps to collect info from your browsers or devices. Their use of cookies and other data gathering tech follows their own privacy rules, not ours.

## Evaluation

This provision governs Yahoo's information sharing, an important aspect of today's markets and economies.<sup>66</sup> The specific challenge here is that the original has inconsistent logic. For example, it says that personally identifiable information about customers will not be shared without consent unless it is with advertisers with whom personally identifiable information can be shared even without consent.

Overall, the clause attempts to set rules that allow Yahoo to share all information in various ways and personally identifiable information in limited ways. We find that the simplification captures this intended logic, although even the best simplification may not overcome the difficulty of parsing what it means for Yahoo to allow for cookie matching with partners without letting them access their cookies. This example provides a more general, important (yet straightforward) insight: AI simplification has limited value if the original legal documents are not properly drafted. Slightly restated, firms might circumvent smart readers and undermine their potential to assist consumers by adopting specific drafting strategies.<sup>67</sup>

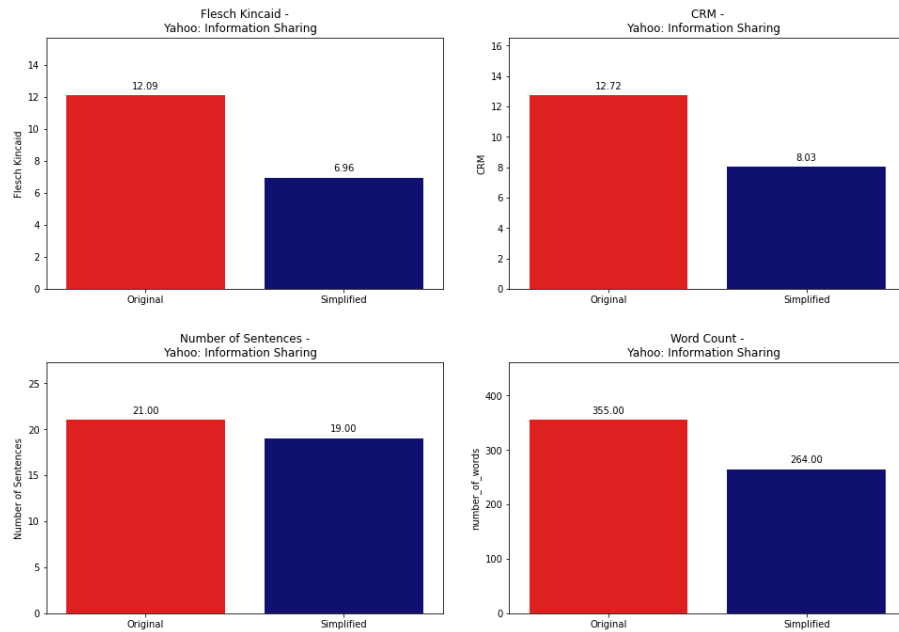
That aside, the simplification is overall helpful. The "you say Okay" is a rather nice touch on "Your consent," although given the quality of consent required (a checkbox), it may overstate the necessary level of consent. It also omits the confusing cookie-matching policy noted above.

More generally, the reduction in text complexity was quite dramatic. The average readability score fell from post-high school level to a level of between seven and eight grade. The length fell by 26% (from 355 words to 264), with a small change to the number of sentences (from 21 to 19). Average word length fell from 5.45 to 4.53 characters per word.

---

<sup>66</sup> Indeed, many scholars have discussed the importance of information sharing. *See, e.g.*, Stacy-Ann Elvy, *Paying for Privacy and the Personal Data Economy*, 117 COL. L. REV. 1369 (2017) (discussing the importance of information sharing); Anja Lambrecht, Avi Goldfarb, Alessandro Bonatti, Anindya Ghose, Daniel G. Goldstein, Randall Lewis, Anita Rao, Navdeep Sahni, & Song Yao, *How Do Firms Make Money Selling Digital Goods Online?*, 25 MKTG. LETTERS 331 (2014) (same); Shmuel I. Becher & Tal Z. Zarsky, *Big Mistake(s)* (Jan. 26, 2024) (unpublished manuscript) (on file with authors) (same).

<sup>67</sup> *See* Arbel & Becher, *supra* note 21, at 118, 141 (discussing adversarial attacks).

**Figure 12: Clause Simplification, Yahoo**

#### H. Spotify: Liability Limitation and Claim Filing

##### Original<sup>68</sup>

##### Limitation of Liability and Time for Filing a Claim

Subject to applicable law, you agree that your sole and exclusive remedy for any problems or dissatisfaction with the Spotify Service is to uninstall any Spotify software and to stop using the Spotify Service. You agree that Spotify has no obligation or liability arising from or related to third-party applications or the content thereof made available through or in connection with the Spotify Service, and while your relationship with such third-party applications may be governed by separate agreements with such third parties, your sole and exclusive remedy, as with respect to Spotify, for any problems or dissatisfaction with any third-party applications or the content thereof, is to uninstall or stop using such third-party applications.

In no event will Spotify, its officers, shareholders, employees, agents, directors, subsidiaries, affiliates, successors, assigns, suppliers, or licensors be liable for (1) any indirect, special, incidental, punitive, exemplary, or consequential damages; (2) any loss of use, data, business, or profits (whether direct or indirect), in all cases arising out of the use of or inability to use the Spotify Service, devices, third-party applications, or third-party application content; or (3) aggregate liability for all claims relating to the Spotify Service, third-party applications, or third-party application content more than the greater of (a) the amounts paid by you to Spotify during the twelve months prior to the first claim; or (b) \$30.00. Any

<sup>68</sup> Spotify Terms, *supra* note 33.

liability we do have for losses you suffer is strictly limited to losses that were reasonably foreseeable.

For clarification, these terms do not limit Spotify's liability for fraud, fraudulent misrepresentation, death or personal injury to the extent that applicable law would prohibit such a limitation and for any other liability that, by applicable law, may not be limited or excluded.

### Simplification

Based on the law, you agree that the only way to fix any issue or unhappiness with the Spotify Service is to remove any Spotify software and stop using the Spotify Service. You agree that Spotify is not to blame or liable for any issues related to apps from other parties or the content from them available through or linked to the Spotify Service. Even though you might have separate deals with these other parties, your only way to fix any issue or unhappiness with any apps from other parties or the content from them, as far as Spotify is concerned, is to remove or stop using such apps from other parties.

Spotify, its staff, shareholders, workers, agents, leaders, child companies, linked companies, future owners, people who take over, suppliers, or those who have a right to give licenses won't be liable for (1) any indirect, special, extra, punitive, exemplary, or follow-on damages; (2) any loss of use, data, business, or money (direct or indirect), in all cases from using or not being able to use the Spotify Service, devices, apps from other parties, or content from apps from other parties; or (3) total liability for all claims about the Spotify Service, apps from other parties, or content from apps from other parties more than the higher of (a) the amounts you paid Spotify in the twelve months before the first claim; or (b) \$30.00. Any liability we do have for losses you go through is strictly limited to losses that could have been reasonably seen coming.

Just to be clear, these terms do not limit Spotify's liability for fraud, wrong representation, death or personal injury to the extent that the law would stop such a limit and for any other liability that, by the law, may not be limited or taken away.

### Evaluation

Liability limitation is a recurring and important theme in consumer contracts, and many jurisdictions treat broad liability limitation clauses with suspicion.<sup>69</sup> Unsurprisingly, Spotify's clause considerably limits its liability. Nonetheless, its wording is inconsistent and confusing. The clause states that the only remedy available to the user is the uninstallation of the app, then moves to limit monetary

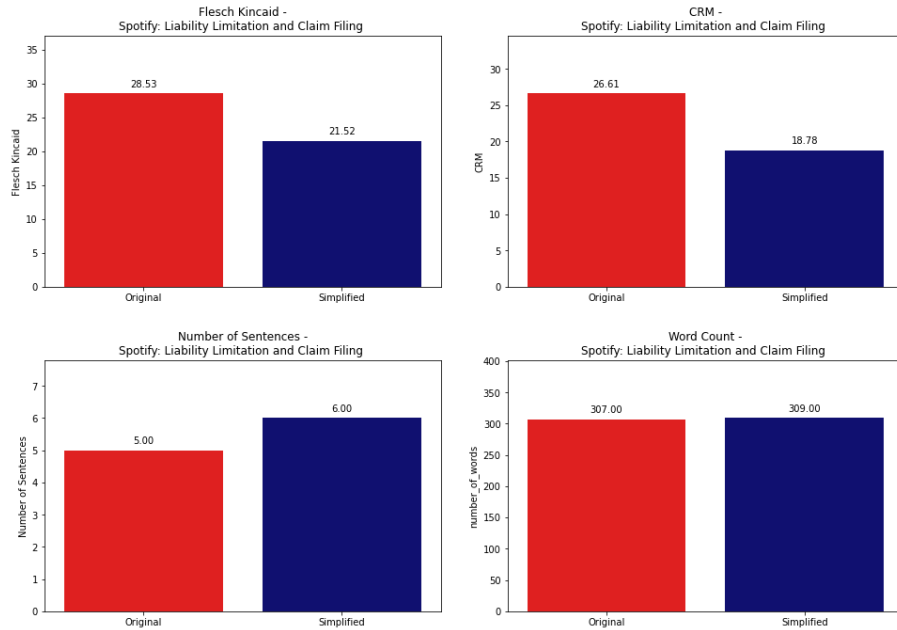
<sup>69</sup> See, e.g., Council Directive 93/13/EEC, of the European Parliament and of the Council of 5 April 1993 on Unfair Terms in Consumer Contracts, annex, 1990 O.J. (L 95) 29. (paragraph 1b of the annex refers to "inappropriately excluding or limiting the legal rights of the consumer vis-à-vis the seller or supplier or another party in the event of total or partial non-performance or inadequate performance by the seller or supplier of any of the contractual obligations...").

remedies to a maximum of \$30, and then states that Spotify *is* liable for anything it cannot disclaim by law.

The summary does an overall good job. It delivers the message of the disclaimer in simple language and references the customer to dispute issues with third parties directly. The emphasis that the liability of a third party is limited “as far as Spotify is concerned” is useful in another aspect: It communicates a certain indifference to such harm, and some consumers may find this warning useful.

However, there are a few noteworthy issues with the simplification. Perhaps guided again by our prompt, the simplification replaces special terms of art. Thus, it changes “incidental damages” to “extra damages,” “consequential damages” to “follow-on damages,” and “foreseeable damages” to “could have seen coming.” Whereas the simplified version might be better in terms of communication, it unduly alters the legal meaning of the clause.

Concerning quantitative metrics, this simplification was the least successful. While it improved readability by seven to eight grade levels, it retained the original structure of the paragraph-long sentence. Readability tests are not well-calibrated to deal with such cases, and the absurd result (28.5 years of schooling exceeds that of almost every lawyer), should be interpreted qualitatively. Furthermore, the simplification made the document marginally longer (from 307 words to 309), suggesting that, at times, text simplification sometimes necessitates more exposition. While there was a slight increase in the number of sentences (from 5 to 6), we observed a reduction in average word length, falling from 5.17 character per word to 4.45.

**Figure 13: Clause Simplification, Spotify**

## VI. SIMPLIFICATION OF SPECIFIC CLAUSES: DISCUSSION

Our analysis emphasizes the role of two aspects of the simplification process: (a) enhanced accessibility (shorter length, reduced complexity, and increased readability) and (b) quality of simplification. Regarding the former, we note that the simplified clauses did well on all the quantitative metrics. In each case, the improvement in readability metrics was significant; on average, reducing the reader's required education level by half.

While we do not necessarily endorse a literal interpretation of the specific grade level assignment, we do recognize the difference as large and meaningful. At the same time, as the Spotify example illustrates, simplified texts might still be inaccessible by standard metrics. The same applies to length and complexity: While the simplified versions presented a substantial improvement on these dimensions, there is no guarantee that consumers are willing to read even these shorter and less complicated texts.

Importantly, the improvement in readability was associated with shorter clauses. This facet is noteworthy, as simplification often requires more exposition. At the same time, the impact on the number of sentences varied, and the effect on word length was relatively mild. However, setting objective metrics aside, we observed an overall significant improvement. The simplified texts were

more accessible and used simple and direct language, making them easier to read and understand.

In terms of accuracy, our report is more mixed. Simplifications of specific clauses were generally accurate and beneficial, especially considering we intentionally chose problematic clauses. Nevertheless, we noticed a range of issues. Some were relatively minor, such as not using a formal tone or failing to include an example that some consumers would find useful. Other issues were more substantial, such as suggesting that consent will be actively sought from the user when it would not, or when omitting the customer's right to complain to state and federal agencies.

An additional significant issue we observed is the incorrect usage of legal terminology. Using "follow-on damages" instead of "consequential damages" entailed more than just confusing the informed consumer. These two concepts of damages are distinct, having very different legal implications. Admittedly, this error may be our own making: the prompt insisted that complex words should be simplified. A less stringent approach would not invite such losses in meaning, although it could compromise the simplification's effectiveness. Yet, to know whether a given term is a term-of-art or a colloquial term requires some domain expertise, and this may point to a limitation of at least current generation general models.

This last point introduces a potentially thorny issue. We generally assume that the canonical contract is the one held by the seller, rather than the one interpreted by the smart reader. However, we wonder if some sellers will persuade the court to adopt the smart reader's version when it serves them by arguing that this is the version the customer presumably used. If courts follow this path, changes to terms-of-art can be harmful to the buyer.

Finally, our analysis did not touch on bias, toxicity, and hallucinations—several issues that afflict current generation models.<sup>70</sup> These issues were fairly muted in our analysis, but we do expect them to become relevant as consumers use smart readers more frequently. We thus acknowledge that our inspection is limited in ways that future research may seek to tackle.

---

<sup>70</sup> Kathy Baxter & Yoav Schlesinger, *Managing the Risks of Generative AI*, HARV. BUS. REV. (June 6, 2023), <https://hbr.org/2023/06/managing-the-risks-of-generative-ai> [<https://perma.cc/2Q7Y-T86Q>].

## VII. SUMMARY

Our study suggests that smart readers can substantially shorten legal texts, reduce their complexity, and improve their readability. Our assessment also indicates that smart readers typically identify important information and include this information in their summaries. Overall, it was encouraging and impressive to find that the simplifications did not substantially undermine the quality of the text and the scope of information that consumers receive.<sup>71</sup> Thus, if consumers choose to use smart readers, this decision could have significant impact on the viability of HIDE strategies and other market outcomes.

The in-depth evaluation of the simplification of specific clauses reinforced our conclusion that while the overall quality of simplification is very high, it is not perfect. However, perfection is not the benchmark. Considerable literature finds that in many domains, consumers rarely read dense legal texts. In such cases, consumers proceed with a vague understanding of the underlying transaction. Provided that smart readers simplify contracts and policies and make them readable, then as long as they are not materially misleading, they can enhance consumers' decision-making. Specifically, smart readers could potentially facilitate informed decisions, enhance efficiency, and thus encourage competition over terms and pressure sellers to draft better contracts. Furthermore, many consumers may wish to examine legal texts *ex post*, once or a dispute arises or when encountering an issue with the transaction they entered. Smart readers can serve well these consumers, who are likely to have more tailored queries and specific aspects to decipher.

Ultimately, simplification tools do not replace lawyers and do not render the original drafting entirely irrelevant to the end consumer. Still, they offer a marked improvement over current consumers' realistic alternatives, such as not reading the text and misperceiving legal aspects. Furthermore, our analysis suggests that smart readers may not only serve individual consumers, but also

---

<sup>71</sup> It is worth noting that while our quality assessment was labor-intensive and had subjective aspects, future developments may advance automated tools to evaluate the quality of legal summaries at scale. *See, e.g.,* Bianca Steffes, Piotr Rataj, Luise Burger, & Lukas Roth, *On Evaluating Legal Summaries With ROUGE*, in PROC. NINETEENTH INT'L CONF. A.I. & L. 475 (2023) (finding that current tools are insufficient for quality evaluation and suggesting to increase the reliability of ROUGE by pre-selecting sentences); *see also* Huyen Nguyen & Junhua Ding, *Keyword-Based Augmentation Method to Enhance Abstractive Summarization for Legal Documents*, in PROC. NINETEENTH INT'L CONF. A.I. & L. 437 (2023) (finding that keywords-based augmentation is effective in improving quality and enhancing summarization models).

energize intermediaries and consumer organizations who can scale up the use of such tools.

Notably, we used non-specialized models with little special domain training to deliver these results, which implies that our findings represent a lower bound, rather than an upper limit. It seems more realistic than ever that smart readers will soon have the ability to automatically detect problematic terms, warn consumers about them, evaluate contracts on a scale, compare contracts, and benchmark them.<sup>72</sup> Consumers can consult with a smart reader in a Q&A mode, asking questions like “What happens if I don’t pay the balance in full?” or “Can I switch providers when I want to?”. Such advances can transform consumer contracting.

Against this revolutionary potential, it is crucial to keep the concerns around accuracy, capture, and bias in mind. Today’s models make errors, and these errors may not be neutral.<sup>73</sup> It is possible that as smart readers grow in influence, companies will wish to influence their output or find ways to mislead them or circumvent their potential. It is, therefore, necessary to consider these concerns when evaluating the technology and its trajectory. In this context, massive open-source models may limit the potential for invisible model corruption.

Our assessment of current generation models concludes that smart readers have arrived. They are not (yet) a full replacement for careful review by a lawyer. However, for the large mass of contracts and privacy policies that today go unread, they serve as a cheap, effective, and scalable alternative. If their potential materializes, a law and policy paradigm shift would be appropriate, if not inevitable.

---

<sup>72</sup> Combining language models with technologies that can automate the detection of unfair terms in consumer contracts is a promising path to consider. For a study that experimentally examines the use of machine learning platforms to perform such detection, see Marco Lippi, Przemyslaw Palka, Giuseppe Contissa, Francesca Lagioia, Hans-Wolfgang Micklitz, Yannis Panagis, Giovanni Sartor, & Paolo Torroni, *Automated Detection of Unfair Clauses in Online Consumer Contracts*, 302 LEGAL KNOWLEDGE & INFO. SYS. 145 (2017); see also Arbel & Becher, *supra* note 21, at 106-108 (discussing smart readers and benchmarking).

<sup>73</sup> Kolt, *supra* note 32.