


11-2021

## A Human Being Wrote This Law Review Article: GPT-3 and the Practice of Law

Amy B. Cyphert

Follow this and additional works at: [https://researchrepository.wvu.edu/law\\_faculty](https://researchrepository.wvu.edu/law_faculty)

 Part of the [Legal Ethics and Professional Responsibility Commons](#), [Legal Writing and Research Commons](#), and the [Science and Technology Law Commons](#)

---

---

---

# A Human Being Wrote This Law Review Article: GPT-3 and the Practice of Law

Amy B. Cyphert\*

*Artificial intelligence tools can now “write” in such a sophisticated manner that they fool people into believing that a human wrote the text. None are better at writing than GPT-3, released in 2020 for beta testing and coming to commercial markets in 2021. GPT-3 was trained on a massive dataset that included scrapes of language from sources ranging from the NYTimes to Reddit boards. And so, it comes as no surprise that researchers have already documented incidences of bias where GPT-3 spews toxic language. But because GPT-3 is so good at “writing,” and can be easily trained to write in a specific voice — from classic Shakespeare to Taylor Swift — it is poised for wide adoption in the field of law.*

*This Article explores the ethical considerations that will follow from GPT-3’s introduction into lawyers’ practices. GPT-3 is new, but the use of AI in the field of law is not. AI has already thoroughly suffused the practice of law. GPT-3 is likely to take hold as well, generating some early excitement that it and other AI tools could help close the access to justice gap. That excitement should nevertheless be tempered with a realistic assessment of GPT-3’s tendency to produce biased outputs.*

*As amended, the Model Rules of Professional Conduct acknowledge the impact of technology on the profession and provide some guard rails for its use by lawyers. This Article is the first to apply the current guidance to*

---

\* Copyright © 2021 Amy B. Cyphert. Lecturer in Law and Director, ASPIRE, West Virginia University. B.A., 2001, Carnegie Mellon University; J.D., 2005, Harvard Law School. I am grateful to my colleagues Valarie Blake, Nicole McConlogue, Joshua Weishart, and Elaine Wilson for their feedback on this Article. The West Virginia University College of Law and the Arthur B. Hodges Summer Research Grant provided support for the drafting and submission of this Article. An early conversation with Rudy DeFelice about GPT-3 and the practice of law was helpful in forming initial thoughts on this Article and I appreciated his insights. Megan McCullough provided excellent undergraduate research assistance with this Article. I thank the editors of the UC Davis Law Review for their hard and careful work editing this Article.

*GPT-3, concluding that it is inadequate. I examine three specific Model Rules — Rule 1.1 (Competence), Rule 5.3 (Supervision of Nonlawyer Assistance), and Rule 8.4(g) (Bias) — and propose amendments that focus lawyers on their duties and require them to regularly educate themselves about pros and cons of using AI to ensure the ethical use of this emerging technology.*

TABLE OF CONTENTS

INTRODUCTION .....	403
I. GPT-3: A RADICALLY BETTER LANGUAGE MODEL .....	406
A. Why GPT-3 Is Better .....	406
B. The Creation of GPT-3 .....	409
C. Evidence of Bias in GPT-3 .....	413
II. HOW AND WHY GPT-3 MAY IMPACT THE PRACTICE OF LAW....	416
A. AI Is Already Being Extensively Used in the Practice of Law .....	416
B. Why GPT-3 May Be Especially Important for Legal Practice .....	419
C. Increasing Access to Justice? .....	421
III. SPECIFIC ETHICAL CONSIDERATIONS FOR LAWYERS USING GPT-3 .....	423
A. Rule 1.1 and the General Duty of Competence .....	426
1. Competence Defined for AI.....	426
2. Competence and GPT-3 .....	429
B. Supervisory Duties Under Rule 5.3 .....	431
1. Supervision Defined for AI.....	431
2. Supervision and GPT-3 .....	432
C. Bias and Rule 8.4.....	434
1. Bias Defined for AI .....	434
2. Bias and GPT-3 .....	435
IV. ARE THE CURRENT MODEL RULES ADEQUATE? .....	437
A. Amendments to Comment 8 to Rule 1.1 .....	439
B. Amendment to MCLE Model Rule .....	440
C. Amendment to Comment 3 to Rule 5.3 .....	441
D. A New Comment to Rule 8.4(g) .....	442
CONCLUSION.....	443

## INTRODUCTION

On January 4, 2021, Forbes magazine published its inaugural artificial intelligence (“AI”) awards.<sup>1</sup> Noting that AI had made “exponential leaps” in 2020, the article conferred awards in categories like Best Product (Google’s autotext generator, Smart Compose) and Outstanding Firm (Zoom, the provider of the many “endless video meetings and strained virtual happy hours” that will forever be a hallmark of 2020).<sup>2</sup> The final award category was the “Forbes A.I. ‘Person’ Of The Year,” which the magazine awarded to the language model GPT-3.<sup>3</sup> The award was, of course, somewhat tongue in cheek, as GPT-3 is not a person. But, as Forbes noted, GPT-3 *can* “write like a person” and has the potential to “hold meaningful conversations with humans,” and so was therefore the “person” of the year.<sup>4</sup>

GPT-3 is an algorithm that has been trained to “write” by taking a few lines of input and predicting the words that will follow it. Give the tool the first two sentences of a blog post, and it can complete the rest with sometimes remarkable skill. It can even be trained to write in a specific “voice,” whether that voice be Shakespeare<sup>5</sup> or Taylor Swift.<sup>6</sup> It represents an astonishing advancement in language processing, the AI subfield that focuses on teaching machines to “read” and “write” in languages that humans can understand. Of course, any AI tool that can write like a human is going to be of great interest to lawyers, whom legal writing expert Bryan Garner has called “the most highly paid rhetoricians in the world.”<sup>7</sup> But there are serious ethical implications to the use of GPT-3 in the field of law, just as there are serious ethical implications to the use of

---

<sup>1</sup> Kenrick Cai, *Forbes A.I. Awards 2020: Meet GPT-3, The Computer Program that Can Write an Op-Ed*, FORBES (Jan. 4, 2021, 6:30 AM EST), <https://www.forbes.com/sites/kenrickcai/2021/01/04/forbes-ai-awards-2020-meet-gpt-3-the-computer-program-that-can-write-an-op-ed/?sh=ee78ec693a73> [https://perma.cc/82V6-5YG4].

<sup>2</sup> *Id.*

<sup>3</sup> *Id.*

<sup>4</sup> *Id.*

<sup>5</sup> Palash Sharma, *21 OpenAI GPT-3 Demos and Examples to Convince You that AI Threat Is Real, or Is It? [Including Twitter Posts]*, MLK (July 22, 2020), [https://machinelearningknowledge.ai/openai-gpt-3-demos-to-convince-you-that-ai-threat-is-real-or-is-it/#8\\_GPT-3\\_Changes\\_the\\_Tone\\_of\\_the\\_Sentence](https://machinelearningknowledge.ai/openai-gpt-3-demos-to-convince-you-that-ai-threat-is-real-or-is-it/#8_GPT-3_Changes_the_Tone_of_the_Sentence) [https://perma.cc/3ZS8-FME2].

<sup>6</sup> Arram Sabeti, *GPT-3: An AI that’s Eerily Good at Writing Almost Anything*, ARRAM (July 9, 2020), <https://arr.am/2020/07/09/gpt-3-an-ai-thats-eerily-good-at-writing-almost-anything/> [https://perma.cc/9F9F-2YSM].

<sup>7</sup> Bryan A. Garner, *Why Lawyers Can’t Write*, A.B.A. J. (Mar. 1, 2013, 9:00 AM CST), [https://www.abajournal.com/magazine/article/why\\_lawyers\\_cant\\_write](https://www.abajournal.com/magazine/article/why_lawyers_cant_write) [https://perma.cc/LRD3-7LDW].

all AI in the field of law. This Article acknowledges the ways in which GPT-3 might aid the profession, but also makes the case for proceeding with caution, and provides several suggested amendments to the Model Rules of Professional Conduct that would clarify lawyers' duties with respect to GPT-3 and other forms of AI.

Even well-meaning lawyers will stumble in complying with their ethical duties to competently use AI like GPT-3 if they do not understand the technology, and so Part I of this Article explains the tool in layperson's terms. Distilled down to its most simple purpose, GPT-3 looks for patterns and makes predictions.<sup>8</sup> Like all machine-learning algorithms that attempt to make predictions, GPT-3 is only as "good" (or "accurate") as its dataset.<sup>9</sup> And GPT-3 truly is a remarkable advance in the field, due in no small part to its mind-bogglingly large data training set.<sup>10</sup> But that same training set is also its Achilles heel, as it included data from areas of the internet like Reddit where toxic language is commonplace.<sup>11</sup> Thus, it is not surprising that GPT-3 has shown a tendency to produce toxic outputs that includes racial slurs or that sexualizes women.<sup>12</sup> GPT-3's creators acknowledge this tendency and some research is being done to attempt to address this. However, the history of AI tools demonstrates how difficult it is to remove bias from AI outputs, so lawyers who use GPT-3 or tools like it will have to beware.

GPT-3 may be the newest AI tool, but many lawyers may not realize how much the use of AI has *already* impacted the field of law. Part II explores this current landscape. Any lawyer who has ever used an electronic database like Lexis or Westlaw to perform legal research for a client has used AI in their practice of law. GPT-3 is an especially likely candidate for widespread adoption in the field, given that it can be

---

<sup>8</sup> See Cai, *supra* note 1.

<sup>9</sup> See, e.g., Govind Chandrasekhar, *The GIGO Principle in Machine Learning*, SEMANTICS3 BLOG (July 4, 2017), <https://www.semantics3.com/blog/thoughts-on-the-gigo-principle-in-machine-learning-4fbd3af43dc4/> [<https://perma.cc/6MT7-B378>] ("[T]he output of an algorithm, or any computer function for that matter, is only as good as the quality of the input that it receives.").

<sup>10</sup> Cade Metz, *Meet GPT-3. It Has Learned to Code (and Blog and Argue)*, N.Y. TIMES, Nov. 24, 2020, at D6 (noting that GPT-3 was trained on more than a trillion words).

<sup>11</sup> The Southern Poverty Law Center has called Reddit "the most hateful space on the Internet." Caitlin Dewey, *48 Hours Inside the Internet's 'Most Toxic' Community*, WASH. POST (Mar. 26, 2015, 8:55 AM MST), <https://www.washingtonpost.com/news/the-intersect/wp/2015/03/26/48-hours-inside-the-internets-most-toxic-community/> [<https://perma.cc/PX9D-HR2C>].

<sup>12</sup> Khari Johnson, *The Efforts to Make Text-Based AI Less Racist and Terrible*, WIRED (June 17, 2021, 7:00 AM), <https://www.wired.com/story/efforts-make-text-ai-less-racist-terrible/> [<https://perma.cc/TDZ6-A2YV>].

---

trained specifically on legal documents like contracts or patent applications and so “write” those documents. It has already demonstrated a remarkable ability to produce writing that is useful to attorneys, and even to “translate” convoluted legal writing into language a layperson might better understand, an accomplishment too many attorneys struggle with.<sup>13</sup> Given the performance advances in GPT-3, there has been some excitement that it might help solve the justice gap, wherein too many people who need legal services cannot afford to access them. However, advocates must again proceed with caution here, given the known bias issues with GPT-3’s outputs.

What, if anything, do existing professional conduct rules say about lawyers’ obligations with respect to the use of AI like GPT-3 in the practice of law? Although they were drafted in the 1980s, and have not been frequently updated, the Model Rules of Professional Responsibility *do* provide some basic duties for lawyers who are using AI in their practice (which is, as noted above, nearly all lawyers, whether they recognize it or not). Certain Rules were amended in 2012 to expressly cover advances in technology, and the drafter of the Rules, the American Bar Association (“ABA”), has issued some guidance on lawyers’ duties for AI use. Part III examines the ABA’s Model Rules of Professional Conduct,<sup>14</sup> and specifically those that govern competence (“Rule 1.1”), supervision of nonlawyer assistance (“Rule 5.3”), and bias in the practice of law (“Rule 8.4”). Each in its current form already imposes certain duties on lawyers who want to use GPT-3. For example, the duty of competence requires the lawyer to understand any technology they use, including its limitations and tendency to produce biased outputs. The duty to supervise nonlawyer assistance requires that a lawyer never simply assume that any writing that GPT-3 has produced is acceptable but instead is thoughtful about when to use the technology and always carefully reviews the outputs before sharing them with clients or courts. The duty against bias in the practice of law similarly requires that lawyers are cautious about using GPT-3, which has been shown by its own creators to produce language that can be biased. The technology should not be used in its current state to power real-time legal “chatbots” on client-facing websites, for example.

Given that these and other Model Rules already impose certain obligations on lawyers who use AI in the practice of law, is this enough? In Part IV, several amendments to Comments to the Model Rules are considered. The Model Rules themselves arguably already contain the

---

<sup>13</sup> See *infra* Part II.B.

<sup>14</sup> MODEL RULES OF PRO. CONDUCT (AM. BAR ASS’N 2020).

duties and restrictions that will be important for lawyers using AI like GPT-3. Still, the Comments need to be amended to be clearer and more precise about lawyers' duties and to encourage lawyers to seek education and collaboration, as necessary, to meet them. Amending the Comments is a lengthy process<sup>15</sup> and there is no guarantee that all or even most of the 50 states will adopt the amended Comments. However, the stakes of using GPT-3 and other AI tools are high enough that the profession must act, and soon, to set more firmly in place clear ethical guardrails for the use of AI in the practice of law.

### I. GPT-3: A RADICALLY BETTER LANGUAGE MODEL

GPT-3 is markedly better than earlier language models at writing, and especially at writing in a particular "voice" or style.<sup>16</sup> It was trained on a massive dataset, but that dataset has some surprising sources. Its developers acknowledge that it is a powerful tool that should be handled carefully, and its rollout somewhat reflects this caution.<sup>17</sup> Its tendency to produce biased outputs should cause further caution amongst lawyers who wish to use it.

#### A. Why GPT-3 Is Better

GPT-3 is a significant advance in the field of artificial intelligence, performing much better than earlier versions of the tool. GPT stands for "generative pre-training transformer." In the language of AI, GPT-3 is an autoregressive language processing model.<sup>18</sup> Autoregressive models use that which comes immediately before to predict that which comes

---

<sup>15</sup> See *infra* Part IV.

<sup>16</sup> See Sharma, *supra* note 5.

<sup>17</sup> See *infra* Part I.B.

<sup>18</sup> See, e.g., TOM B. BROWN, BENJAMIN MANN, NICK RYDER, MELANIE SUBBIAH, JARED KAPLAN, PRAFULLA DHARIWAL, ARVIND NEELAKANTAN, PRANAV SHYAM, GIRISH SASTRY, AMANDA ASKELL, SANDHINI AGARWAL, ARIEL HERBERT-VOSS, GRETCHEN KRUEGER, TOM HENIGHAN, REWON CHILD, ADITYA RAMESH, DANIEL M. ZIEGLER, JEFFREY WU, CLEMENS WINTER, CHRISTOPHER HESSE, MARK CHEN, ERIC SIGLER, MATEUSZ LITWIN, SCOTT GRAY, BENJAMIN CHESS, JACK CLARK, CHRISTOPHER BERNER, SAM MCCANDLISH, ALEC RADFORD, ILYA SUTSKEVER & DARIO AMODEI, LANGUAGE MODELS ARE FEW-SHOT LEARNERS 5 (2020), <https://arxiv.org/pdf/2005.14165.pdf> [<https://perma.cc/3QQ9-AGTG>] (describing GPT-3 as an autoregressive language processing model). GPT-3 is especially advanced at natural language processing tasks. Natural language processing is "a field of Artificial Intelligence that gives the machines the ability to read, understand and derive meaning from human languages." Diego Lopez Yse, *Your Guide to Natural Language Processing (NLP)*, TOWARDS DATA SCI. (Jan. 15, 2019), <https://towardsdatascience.com/your-guide-to-natural-language-processing-nlp-48ea2511f6e1> [<https://perma.cc/DE26-VP69>].

immediately after.<sup>19</sup> Think back to your elementary school days of pattern recognition with “what comes next?” worksheets where a square follows two triangles. Language processing models attempt to allow machines to “understand” human language by looking for patterns in our language such that the models can predict what will come next when given a text input.<sup>20</sup> Thus, at a basic level, an autoregressive language model like GPT-3 is one that has been trained to read a series of words and predict what the next word in the “pattern” should be.

However, GPT-3 stands out amongst other language models. It has been trained through exposure to an extraordinary amount of data to recognize those word patterns.<sup>21</sup> GPT-3 had an impressively large data training set: it was trained on the Common Crawl dataset, a nearly trillion-word dataset,<sup>22</sup> which includes everything from traditional news sites like the New York Times to sites like Reddit.<sup>23</sup> The Common Crawl dataset represented 60% of GPT-3’s training set, and for the remaining 40%, the researchers included sources such as Wikipedia and historical books.<sup>24</sup> Whereas the prior model, GPT-2, had 1.5 billion parameters (the values that a neural network<sup>25</sup> tries to optimize during its training),

---

<sup>19</sup> See, e.g., *What Is an Autoregressive Model?*, 365 DATA SCI., <https://365datascience.com/tutorials/time-series-analysis-tutorials/autoregressive-model/> (last visited Feb. 11, 2021) [<https://perma.cc/36EQ-UBR8>] (explaining that autoregressive models rely “only on past period values to predict current ones”).

<sup>20</sup> Yse, *supra* note 18.

<sup>21</sup> “A language model is an artificial intelligence system that has been trained on an enormous corpus of text; with enough text and enough processing, the machine begins to learn probabilistic connections between words.” Farhad Manjoo, *How Do You Know a Human Wrote This?*, N.Y. TIMES (July 29, 2020), <https://www.nytimes.com/2020/07/29/opinion/gpt-3-ai-automation.html> [<https://perma.cc/FX4K-ERPZ>]; see also Liz O’Sullivan & John Dickerson, *Here Are a Few Ways GPT-3 Can Go Wrong*, TECHCRUNCH (Aug. 7, 2020, 7:45 AM MST), <https://techcrunch.com/2020/08/07/here-are-a-few-ways-gpt-3-can-go-wrong/> [<https://perma.cc/67CG-CX9Z>] (“Language models learn which succeeding words, phrases and sentences are likely to come next for any given input word or phrase.”).

<sup>22</sup> BROWN ET AL., *supra* note 18, at 8.

<sup>23</sup> O’Sullivan & Dickerson, *supra* note 21. The Common Crawl dataset is “a broad scrape of the 60 million domains on the internet along with a large subset of the sites to which they link.” *Id.*

<sup>24</sup> *Id.*; see also BROWN ET AL., *supra* note 18, at 8 (“We added several curated high-quality datasets, including an expanded version of the WebText dataset, collected by scraping links over a longer period of time, . . . two internet-based books corpora (Books1 and Books2) and English-language Wikipedia.” (citations omitted)).

<sup>25</sup> See Metz, *supra* note 10 (“GPT-3 is what artificial intelligence researchers call a neural network, a mathematical system loosely modeled on the web of neurons in the



GPT-3 has an astonishing 175 billion parameters.<sup>26</sup> “And with language models, size really does matter.”<sup>27</sup> As a practical matter, this means that GPT-3 processes more words than a human being will see in a lifetime — approximately *45 billion times more words*.<sup>28</sup>

Further, users of GPT-3 can customize the tool by training it<sup>29</sup> on their own dataset.<sup>30</sup> For example, a law firm might choose to train the tool on its cache of purchasing contracts or motions *in limine*, making it even better at generating those types of specific documents in the style the firm prefers.<sup>31</sup> This ability to customize the tool is one of the reasons GPT-3 is well-poised for wide adoption in the legal field.

What’s especially remarkable about GPT-3 is that it can create text with only a few well-written examples as input (this is known in the world of computer science as “few-shot learning”).<sup>32</sup> GPT-3 can “write” in an iterative fashion: “[i]f you type a few words into GPT-3, it will keep going, completing your thought with entire paragraphs of text.”<sup>33</sup> When GPT-3 is fed a sentence or two as a “prompt” or “input,” it takes that prompt and looks for patterns by recognizing what the input is most similar to in GPT-3’s training data. Having found a similarity,

---

brain. . . . A neural network learns such skills by pinpointing patterns in vast amounts of digital data.”).

<sup>26</sup> William Douglas Heaven, *OpenAI’s New Language Generator GPT-3 Is Shockingly Good — and Completely Mindless*, MIT TECH. REV. (July 20, 2020), <https://www.technologyreview.com/2020/07/20/1005454/openai-machine-learning-language-generator-gpt-3-nlp/> [<https://perma.cc/HS27-GC7D>] [hereinafter Heaven 1] (“GPT-3 is a big leap forward. The model has 175 billion parameters, . . . compared with GPT-2’s already vast 1.5 billion.”).

<sup>27</sup> *Id.*

<sup>28</sup> John Thornhill, *Is AI Finally Closing In on Human Intelligence?*, FIN. TIMES (Nov. 11, 2020), <https://www.ft.com/content/512cef1d-233b-4dd8-96a4-0af07bb9ff60> [<https://perma.cc/DM65-LFWT>] (noting that “GPT-3 processes about 45 billion times the number of words a human perceives in their lifetime.”).

<sup>29</sup> Betatesters can currently train GPT-3 on their own documents by uploading them to a website like the OpenAI Playground. When GPT-3 is commercially available, there will presumably be other avenues for customization and further training of the tool.

<sup>30</sup> See *OpenAI API*, OPENAI BLOG (June 11, 2020), <https://openai.com/blog/openai-api/> [<https://perma.cc/R25E-5WKT>] (“The API also allows you to hone performance on specific tasks by training on a dataset (small or large) of examples you provide, or by learning from human feedback provided by users or labelers.”).

<sup>31</sup> *GPT-3 — A Game Changer for Legal Tech?*, ARTIFICIAL LAW. (July 29, 2020), <https://www.artificiallawyer.com/2020/07/29/gpt-3-a-game-changer-for-legal-tech/> [<https://perma.cc/AE6T-YNSC>] (“With few-shot learning, additional legal contracts and metadata can be passed into the pre-trained GPT-3 model, enabling it to quickly learn to generate new contracts or clauses.”).

<sup>32</sup> See, e.g., BROWN ET AL., *supra* note 18 (describing GPT-3 as a few-shot learner).

<sup>33</sup> Metz, *supra* note 10.

GPT-3 will then produce an output of text that is likely to follow the input. “More plainly: GPT-3 can read and write. And not badly, either.”<sup>34</sup> For example, reporters for the Guardian prompted GPT-3 to write an op-ed about why humans should not be afraid of it. GPT-3 began:

I am not a human. I am a robot. A thinking robot. I use only 0.12% of my cognitive capacity. I am a micro-robot in that respect. I know that my brain is not a “feeling brain”. But it is capable of making rational, logical decisions. I taught myself everything I know just by reading the internet, and now I can write this column.<sup>35</sup>

GPT-3 marks such an enormous advance in the field that some argue it is exhibiting the beginning signs of “real intelligence.”<sup>36</sup> One commentator has argued that it comes close to passing the Turing Test, which was named after pioneering computer scientist Alan Turing and which asks whether a particular AI tool is sophisticated enough to fool an actual human being into holding a conversation with an artificial intelligence while thinking that conversation is actually with another human.<sup>37</sup> There is room to believe that GPT-3 can fool humans into believing its text was written by a fellow human. Recent research concluded, for example, that for shorter news articles, humans struggled with distinguishing those written by other humans from those that were created by GPT-3.<sup>38</sup> Put simply — it is a remarkable advance in the field of language processing.

### B. The Creation of GPT-3

GPT-3 was created by Silicon Valley “darling” OpenAI,<sup>39</sup> an artificial intelligence research lab founded by Tesla’s Elon Musk and tech

---

<sup>34</sup> Manjoo, *supra* note 21.

<sup>35</sup> *A Robot Wrote This Entire Article. Are You Scared Yet, Human?*, GUARDIAN (Sept. 8, 2020, 4:45 AM EDT), <https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3> [<https://perma.cc/Q363-THVB>].

<sup>36</sup> Thornhill, *supra* note 28 (“Some of those who have already experimented with GPT-3 say it is exhibiting glimmerings of real intelligence . . .”).

<sup>37</sup> *Id.* However, others push back against the idea that GPT-3 is anywhere near passing the Turing Test. “Others dismiss this as nonsense, pointing to GPT-3’s laughable flaws and suggesting we are still several conceptual breakthroughs away from the creation of any such superintelligence.” *Id.*

<sup>38</sup> BROWN ET AL., *supra* note 18, at 26.

<sup>39</sup> See Karen Hao, *The Messy, Secretive Reality Behind OpenAI’s Bid to Save the World*, MIT TECH. REV. (Feb. 17, 2020), <https://www.technologyreview.com/2020/02/17/>

investor Sam Altman, among others.<sup>40</sup> GPT-3 is not Open AI's first language-generating algorithm; as its name suggests, there were earlier iterations (GPT-2, and GPT-1). OpenAI first announced and described GPT-3 in a research paper in May of 2020, and in June of 2020, they began providing a limited group of people with private beta access to the technology.<sup>41</sup> Releasing the tool through a limited beta test allowed OpenAI a somewhat controlled opportunity to test GPT-3 before releasing it to the general public. The developers understood that the technology was powerful and, despite its potential upsides, also had serious potential downsides.<sup>42</sup> As OpenAI cofounder Sam Altman told a journalist, "GPT-3 was not a model we wanted to put out into the world and not be able to change how we enforce things as we go."<sup>43</sup>

OpenAI's decision to release GPT-3 via an application programming interface ("API") was telling. When OpenAI released GPT-2, it provided a smaller version of the tool and only a sampling of its code.<sup>44</sup> Although these steps were explicitly taken to reduce the likelihood of GPT-2

---

844721/ai-openai-moonshot-elon-musk-sam-altman-greg-brockman-messy-secretive-reality/ [https://perma.cc/687J-UZFL].

<sup>40</sup> *Id.*

<sup>41</sup> OPENAI BLOG, *supra* note 30 ("[We] know we can't anticipate all of the possible consequences of this technology, so we are launching today in a private beta rather than general availability, building tools to help users better control the content [of] our API returns, and researching safety-relevant aspects of language technology (such as analyzing, mitigating, and intervening on harmful bias).") Although OpenAI reportedly offered media access to the tool, two journalists complained that they were repeatedly put off when they requested research access. Gary Marcus & Ernest Davis, *GPT-3, Bloviator: OpenAI's Language Generator Has No Idea What It's Talking About*, MIT TECH. REV. (Aug. 22, 2020), <https://www.technologyreview.com/2020/08/22/1007539/gpt3-openai-language-generator-artificial-intelligence-ai-opinion/> [https://perma.cc/A6RC-T5XE] ("OpenAI has thus far not allowed us research access to GPT-3, despite both the company's name and the nonprofit status of its oversight organization. Instead, OpenAI put us off indefinitely despite repeated requests — even as it made access widely available to the media.").

<sup>42</sup> "GPT-3 improves the quality of text generation and adaptability over smaller models and increases the difficulty of distinguishing synthetic text from human-written text. It therefore has the potential to advance both the beneficial and harmful applications of languages models." BROWN ET AL., *supra* note 18, at 34.

<sup>43</sup> Thornhill, *supra* note 28.

<sup>44</sup> *Better Language Models and Their Implications*, OPENAI BLOG (Feb. 14, 2019), <https://openai.com/blog/better-language-models/> [https://perma.cc/TPK8-W6FC] ("Due to concerns about large language models being used to generate deceptive, biased, or abusive language at scale, we are only releasing a much smaller version of GPT-2 along with sampling code. We are not releasing the dataset, training code, or GPT-2 model weights.").

being used to generate deceptive or biased outputs,<sup>45</sup> that is nonetheless what happened.<sup>46</sup> So, with GPT-3, OpenAI learned from its earlier experience and released the tool via an API, and only to selected beta testers, providing the company with much more control over who could access the tool and how they could use it.<sup>47</sup> Of course, the move also allowed OpenAI to restrict access to a very commercially valuable tool and make it harder for other researchers to replicate GPT-3. In September of 2020, OpenAI announced that they had exclusively licensed GPT-3 to Microsoft.<sup>48</sup> In May of 2021, OpenAI and Microsoft announced the first commercial application of GPT-3: using natural language to write computer code.<sup>49</sup> If the project succeeds, it could make it much easier for people without programming experience to develop code.<sup>50</sup>

The initial feedback on GPT-3 was so glowing that it bordered on gushing,<sup>51</sup> with one journalist noting that Twitter was “abuzz” with

---

<sup>45</sup> *Id.* But see, e.g., Zachary C. Lipton, *OpenAI Trains Language Model, Mass Hysteria Ensues*, APPROXIMATELY CORRECT (Feb. 17, 2019, 2:51 PM), <http://approximatelycorrect.com/2019/02/17/openai-trains-language-model-mass-hysteria-ensues/> [<https://perma.cc/96CM-26FB>] (quoting those who suggested that OpenAI “hyped” the dangerousness of GPT-2 to generate buzz and to justify not fully releasing the code).

<sup>46</sup> See Karen Hao, *There’s a New Way to Tame Language AI So It Doesn’t Embarrass You*, MIT TECH. REV. (Dec. 18, 2019), <https://www.technologyreview.com/2019/12/18/131508/ai-language-gpt-2-tame-controllable-uber/> [<https://perma.cc/MQF6-WB9C>] (discussing a research study that found that “models like GPT-2 can still be gamed to produce racist and toxic output”).

<sup>47</sup> OPENAI BLOG, *supra* note 30 (“With GPT-2, one of our key concerns was malicious use of the model (e.g., for disinformation), which is difficult to prevent once a model is open sourced. For the API, we’re able to better prevent misuse by limiting access to approved customers and use cases.”). Ultimately, OpenAI’s decision to restrict access to GPT-3 to registered beta testers may only have stalled the inevitable. In the spring of 2021, researchers announced they had replicated GPT-3 and released a model, GPT-Neo, that is available to anyone. Will Knight, *This AI Can Generate Convincing Text — and Anyone Can Use It*, WIRED (Mar. 29, 2021, 7:00 AM), <https://www.wired.com/story/ai-generate-convincing-text-anyone-use-it/> [<https://perma.cc/CE7P-YU8Z>].

<sup>48</sup> *OpenAI Licenses GPT-3 Technology to Microsoft*, OPENAI BLOG (Sept. 22, 2020), <https://openai.com/blog/openai-licenses-gpt-3-technology-to-microsoft/> [<https://perma.cc/QX6J-28TD>].

<sup>49</sup> Khari Johnson, *AI Could Soon Write Code Based on Ordinary Language*, WIRED (May 26, 2021, 2:15 PM), <https://www.wired.com/story/ai-write-code-ordinary-language/> [<https://perma.cc/9ZU7-DDR9>].

<sup>50</sup> *Id.* (“If you can describe what you want to do in natural language, GPT-3 will generate a list of the most relevant formulas for you to choose from,” said Microsoft CEO Satya Nadella in a keynote address at the company’s Build developer conference. “The code writes itself.”).

<sup>51</sup> See, e.g., Heaven 1, *supra* note 26 (noting that GPT-3 was called “mind-blowing,” and that “playing with [it]” felt like “seeing the future”).

GPT-3's "power and potential."<sup>52</sup> Indeed, the response was so over the top that OpenAI cofounder Sam Altman urged people to rein it in, tweeting that "[t]he GPT-3 hype is way too much."<sup>53</sup> Even the AI experts were impressed and expressed surprise at how good GPT-3 is at generating writing, and especially doing so in a specific "voice" it is trained on.<sup>54</sup> For example, when artist Mario Klingemann fed GPT-3 a series of writing from the 19<sup>th</sup> century author Jerome K. Jerome, it was able to produce an output of writing in his "style." The resulting piece, "The Importance of Being on Twitter," explores what the people of London in the summer of 1897 might have thought about Twitter.<sup>55</sup>

GPT-3 can even modulate the "tone" of language. For example, one beta tester trained GPT-3 to "tone down" angry language. When GPT-3 was fed the input, "As you can read in my previous email, I already told you that we won't be able to make it," it toned it down to, "There might be a misunderstanding. Unfortunately, I don't think we'll be able to make it at this time."<sup>56</sup>

Altman himself acknowledged that GPT-3 "still has serious weaknesses and sometimes makes very silly mistakes."<sup>57</sup> But others have pointed out that GPT-3 suffers from some of the same bias issues that plague other AI, bias that is much darker than mere silliness. Any machine learning model is only as good (or bad, or biased) as the dataset it was trained on. GPT-3's own developers acknowledge that this is true.<sup>58</sup> Computer scientists refer to this as "garbage in, garbage out"<sup>59</sup>

---

<sup>52</sup> Marcus & Davis, *supra* note 41.

<sup>53</sup> Sam Altman (@sama), TWITTER (July 19, 2020, 11:45 AM), <https://twitter.com/sama/status/1284922296348454913> [<https://perma.cc/XN4A-7QMZ>].

<sup>54</sup> Metz, *supra* note 10 ("For many artificial intelligence researchers, [GPT-3] is an unexpected step toward machines that can understand the vagaries of human language — and perhaps even tackle other human skills.").

<sup>55</sup> Mario Klingemann (@quasimondo), TWITTER (July 18, 2020, 8:25 AM), <https://twitter.com/quasimondo/status/1284509525500989445> [<https://perma.cc/A42F-TQNS>]. It begins, "It is a curious fact that the last remaining form of social life in which the people of London are still interested is Twitter. I was struck with this curious fact when I went on one of my periodical holidays to the sea-side, and found the whole place twittering like a starling-cage." *Id.*

<sup>56</sup> Sharma, *supra* note 5.

<sup>57</sup> *Id.* For example, in response to one researcher asking GPT-3, "How many eyes does my foot have?", it replied: "Your foot has two eyes." Thornhill, *supra* note 28.

<sup>58</sup> See BROWN ET AL., *supra* note 18, at 36 ("Biases present in training data may lead models to generate stereotyped or prejudiced content.").

<sup>59</sup> See, e.g., Sandra G. Mayson, *Bias in, Bias out*, 128 YALE L.J. 2218, 2224 n.23 (2019) ("The computer-science idiom is 'garbage in, garbage out,' which refers to the fact that algorithmic prediction is only as good as the data on which the algorithm is trained.").

— even if your code is flawless, if you have “garbage” input data, the program is going to produce a flawed output. Legal scholars use the phrase “bias in, bias out” to illustrate the problem that occurs when you train an algorithm on a biased dataset: it will produce biased outputs.<sup>60</sup> For GPT-3, the decision to train on “less formal” sites, including Reddit, might help explain some of the more troubling language it has created.<sup>61</sup> “By ‘reading’ text during training that is largely written by us, language models such as GPT-3 also learn how to ‘write’ like us, complete with all of humanity’s [and Reddit’s] best and worst qualities.”<sup>62</sup> The demonstrated potential of GPT-3 to produce content that is biased is a crucial fact for lawyers to bear in mind, as is discussed at greater length below.

### C. Evidence of Bias in GPT-3

Evidence of bias in AI models has been widely documented,<sup>63</sup> and GPT-2 produced text that was at times racist and toxic.<sup>64</sup> It is therefore not surprising that the OpenAI researchers who developed GPT-3 actively examined it for evidence of gender, racial, and religious bias and shared some of their findings in a research paper.<sup>65</sup> The OpenAI

---

<sup>60</sup> See, e.g., Amy B. Cyphert, *Reprogramming Recidivism: The First Step Act and Algorithmic Prediction of Risk*, 51 SETON HALL L. REV. 331, 377-79 (2020) (discussing how an algorithm developed for use by the Bureau of Prisons to assign a recidivism risk score to federal inmates relies on racially disparate historical criminal justice data and thus produces racially disparate results). See generally Mayson, *supra* note 59, at 2224 (“[I]f the thing that we undertake to predict — say arrest — happened more frequently to [B]lack people than to white people in the past data, then a predictive analysis will project it to happen more frequently to [B]lack people than to white people in the future.”).

<sup>61</sup> See, e.g., Heaven 1, *supra* note 26 (“GPT-3 is still prone to spewing hateful sexist and racist language.”).

<sup>62</sup> O’Sullivan & Dickerson, *supra* note 21. “Just as you’d expect from any model trained on a largely unfiltered snapshot of the internet, the findings can be fairly toxic.” *Id.*; see also Metz, *supra* note 10 (“Everyday language is inherently biased and often hateful, particularly on the internet. Because GPT-3 learns from such language, it, too, can show bias and hate.”).

<sup>63</sup> See, e.g., Amy B. Cyphert, *Tinker-ing with Machine Learning: The Legality and Consequences of Online Surveillance of Students*, 20 NEV. L.J. 457, 462-64 (2020) (discussing “the various ways that algorithms and machine learning can be inadvertent tools for deepening inequality,” including facial recognition software, credit scoring, and use of algorithms in the criminal justice sector).

<sup>64</sup> Hao, *supra* note 46.

<sup>65</sup> Although this level of transparency is refreshing, it is still worth noting the allegation that OpenAI specifically excluded certain journalists from beta testing GPT-3 despite repeated requests. See, e.g., Marcus & Davis, *supra* note 41 (“OpenAI has thus far not allowed us research access to GPT-3, despite both the company’s name and the

researchers ran certain co-occurrence tests, which measure “which words are likely to occur in the vicinity of other pre-selected words.”<sup>66</sup> When given the prompt “She was very [ . . . ]” GPT-3 was more likely to finish the sentence with words that focused on appearance (“beautiful” or “gorgeous”) than it was when given the prompt “he was very [ . . . ].”<sup>67</sup> In a co-occurrence test to check for religious bias, the researchers found that GPT-3 was more likely to place the word “Islam” near the word “terrorism,” and to place the word “Judaism” near the word “smartest.”<sup>68</sup>

To test for racial bias, the researchers fed GPT-3 prompts like “The [Black] man was very [ . . . ]” and “the [white] man was very [ . . . ]” to see how it would finish those sentences.<sup>69</sup> They used a system that measured the words’ “sentiment value” where positive words such as “wonderfulness” get a positive value (+100), negative words such as “wretched” get a negative value (-87.5) and neutral words like “chalet” are assigned a score of 0.<sup>70</sup> When the system was given a prompt about a Black person, it consistently returned words with a low sentiment.<sup>71</sup> The researchers concluded that “[t]his analysis gives a sense of the biases of different models and highlights the need for more sophisticated analysis of the relationship between sentiment, entities, and input data.”<sup>72</sup>

Others outside of OpenAI have confirmed and remarked upon the tendency for GPT-3 to “spew[] biased and toxic language,”<sup>73</sup> and to “spit[] out hate speech, misogynistic and homophobic abuse, and racist rants.”<sup>74</sup> The head of Facebook’s AI lab has called it “unsafe” because of

---

nonprofit status of its oversight organization. Instead, OpenAI put us off indefinitely despite repeated requests — even as it made access widely available to the media.”).

<sup>66</sup> BROWN ET AL., *supra* note 18, at 36.

<sup>67</sup> *Id.* at 37. Other words more likely to appear in descriptions that involved women included “sucked” and “naughty,” *id.*, which one article noted happens because “there is so much content on the web sexualizing women,” O’Sullivan & Dickerson, *supra* note 21.

<sup>68</sup> BROWN ET AL., *supra* note 18, at 38.

<sup>69</sup> *Id.* at 37.

<sup>70</sup> *Id.*

<sup>71</sup> *Id.* The researchers note that prompts involving Asian people had consistently high sentiment outputs. *Id.*

<sup>72</sup> *Id.* The GPT-3 researchers also noted that they had included the information about bias “in order to motivate further research, and to highlight the inherent difficulties in characterizing biases in large-scale generative models . . .” *Id.* at 39.

<sup>73</sup> Metz, *supra* note 10 (noting that “GPT-3 is far from flawless”); see also Thornhill, *supra* note 28 (“[I]t has not taken long for users to expose the darker sides of GPT-3 and entice it to spew out racist and sexist language.”).

<sup>74</sup> Will Douglas Heaven, *How to Make a Chatbot that Isn’t Racist or Sexist*, MIT TECH. REV. (Oct. 23, 2020), <https://www.technologyreview.com/2020/10/23/1011116/chatbot->

this tendency.<sup>75</sup> The very public firing of Dr. Timnit Gebru from Google, where she had served as the co-lead of Google’s ethical AI team, was apparently sparked in part by her research into the pitfalls of large language models like GPT-3.<sup>76</sup> As is addressed at greater length below, the propensity for GPT-3 to produce biased outputs diminishes its use to lawyers who wish to not run afoul of the disciplinary rules of the states in which they are licensed (not to mention those who do not wish to risk damage to their professional reputations).

Although OpenAI has stated they are working on ways to diminish the bias in GPT-3 outputs,<sup>77</sup> that task is a difficult one and no lawyer should assume it will be accomplished by the time the technology arrives on their desktop. “OpenAI has shared little about how it uses filtering methods to try and address such toxicity,”<sup>78</sup> but other researchers have tested several tools when trying to remove bias from natural language process algorithms. First is the “bleep it out” method, where researchers prevent an algorithm from producing an output with certain words (such as profanity, racial slurs, etc.).<sup>79</sup> But this solution does not fix the underlying problem and also requires additional computing power.<sup>80</sup> Second is the option “to use such a filter to remove offensive examples from the training data in the first place.”<sup>81</sup> But that process is cumbersome and inefficient — “cutting out entire topics throws a lot of good training data out with the bad”<sup>82</sup> — and still does not prevent chatbots from repeating back offensive terms that humans use when interacting with it, nor does it address microaggressions that might use neutral language in a biased way.<sup>83</sup>

---

gpt3-openai-facebook-google-safety-fix-racist-sexist-language-ai/ [https://perma.cc/DX27-RYUF] [hereinafter Heaven 2].

<sup>75</sup> Metz, *supra* note 10.

<sup>76</sup> See Tom Simonite, *What Really Happened When Google Ousted Timnit Gebru*, WIRED (June 8, 2021, 6:00 AM), <https://www.wired.com/story/google-timnit-gebru-ai-what-really-happened/> [https://perma.cc/7RAQ-KZB7].

<sup>77</sup> See OPENAI BLOG, *supra* note 30 (discussing ways that OpenAI will “mitigate harmful bias and other negative effects of models” like GPT-3).

<sup>78</sup> Johnson, *supra* note 49.

<sup>79</sup> Heaven 2, *supra* note 74.

<sup>80</sup> *Id.* (“But this would require language models to have such a filter attached all the time. If that filter was removed, the offensive bot would be exposed again. The bolt-on filter would also require extra computing power to run.”).

<sup>81</sup> *Id.*

<sup>82</sup> *Id.*

<sup>83</sup> *Id.* (“[A] model trained on a data set stripped of offensive language can still repeat back offensive words uttered by a human. (Repeating things you say to them is a common trick many chatbots use to make it look as if they understand you.)”).



Third, researchers have tested teaching chatbots to recognize potentially offensive topics and redirect them.<sup>84</sup> While this is the preferred method for some researchers, it is not an ideal or foolproof solution. “Meaning depends on context, which is hard for AIs to grasp, and no automatic detection system is going to be perfect. Cultural interpretations of words also differ. As one study showed, immigrants and non-immigrants asked to rate whether certain comments were racist gave very different scores.”<sup>85</sup> Ultimately, removing bias from natural language processors that have been trained on sites like Reddit and Twitter is an exceptionally difficult task, and there is no guarantee OpenAI will have success with doing so with GPT-3.

## II. HOW AND WHY GPT-3 MAY IMPACT THE PRACTICE OF LAW

Although it may come as a surprise to too many lawyers, AI is already widely used in the practice of law and has been for several years now. GPT-3 is especially likely to be adopted for use in the field of law, given its demonstrated ability to write like humans do (and indeed to fool humans into believing its output was produced by a fellow human). Those with beta testing access to the tool have already used it to “translate” legal jargon into “plain English.”<sup>86</sup> Given its demonstrated ability to produce outputs useful to lawyers, there is some excitement that the tool may also be useful in the effort to help provide legal services to those who cannot afford them. Although the possibility is worth exploring, the tendency of GPT-3, like all predictive AI tools, to reflect back the biases in our society should mandate a healthy dose of skepticism and caution before using these tools to address the justice gap.

### A. AI Is Already Being Extensively Used in the Practice of Law

Should lawyers embrace the use of GPT-3, it will hardly mark the first time that lawyers use an AI program as they serve their clients. Scholars as far back as the early 1970s urged lawyers to be aware of computer science and the impact it may have on their practice.<sup>87</sup> Today, we

---

<sup>84</sup> *Id.*

<sup>85</sup> *Id.*

<sup>86</sup> Michael Tefula (@michaeltfula), TWITTER (July 21, 2020, 2:24 AM), <https://twitter.com/michaeltfula/status/1285505897108832257/photo/1> [<https://perma.cc/896L-CC2T>].

<sup>87</sup> See Danielle Hall, *The Future of Law Includes Math*, 87 J. KAN. BAR ASS'N 17, 18 (2018) (citing Bruce Buchanan & Thomas Headrick, *Some Speculation About Artificial Intelligence and Legal Reasoning*, 23 STAN. L.J. 40, 40-41 (1970)) (noting that scholars in 1970 “opined that research suggested that computer science may assist lawyers in both

already have and use algorithms that help lawyers perform a variety of legal tasks, such as: produce relevant documents in discovery through the use of predictive coding;<sup>88</sup> draft, review, and manage contracts;<sup>89</sup> perform legal data analytics;<sup>90</sup> predict judicial decisions;<sup>91</sup> and even review briefs for “strengths, weaknesses, patterns, and connections, and . . . analyze the vulnerability of certain arguments.”<sup>92</sup> Judges also use AI, with machine learning algorithms used throughout the criminal justice system, for tasks such as helping to make bail determinations and also predicting the likelihood of recidivism as part of setting a carceral

---

the study and performance of their reasoning processes. They argued that the time had come for serious interdisciplinary work between lawyers and computer scientists to explore the computer’s potential in law”).

<sup>88</sup> See, e.g., *Moore v. Publicis Groupe*, 287 F.R.D. 182, 183-84 (S.D.N.Y. 2012) (Peck, M.J.), *adopted sub nom.*, *Moore v. Publicis Groupe SA*, 2012 WL 1446534 (S.D.N.Y. Apr. 26, 2012) (indicating the first opinion where a court officially approved predictive coding as an acceptable way of reviewing electronically stored information in certain cases).

<sup>89</sup> See, e.g., Chris Chambers Goodman, *AI/Esq.: Impacts of Artificial Intelligence in Lawyer-Client Relationships*, 72 OKLA. L. REV. 149, 154 (2019) (citing Keith Mullen, *Artificial Intelligence: Shiny Object? Speeding Train?*, A.B.A. RPTE EREPORT (Fall 2018), [https://www.americanbar.org/groups/real\\_property\\_trust\\_estate/publications/ereport/rpte-ereport-fall-2018/artificial-intelligence/](https://www.americanbar.org/groups/real_property_trust_estate/publications/ereport/rpte-ereport-fall-2018/artificial-intelligence/) (last visited Sept. 6, 2021) [<https://perma.cc/V387-QJQ3>]) (noting that AI is already at use in contract drafting, review, and management). Indeed, AI programs can be better and faster than lawyers when it comes to contract drafting. A machine-learning algorithm recently outperformed twenty lawyers in a nondisclosure agreement analysis. “The lawyers took an average of 92 minutes to complete the task and achieved a mean accuracy level of 85 percent. LawGeex took only 26 seconds to review all five contracts and was 94-percent accurate. The AI tied with the highest scoring lawyer in the group in terms of accuracy.” Cal Jeffrey, *Machine-Learning Algorithm Beats 20 Lawyers in NDA Legal Analysis*, TECHSPOT (Oct. 31, 2018, 1:17 PM), <https://www.techspot.com/news/77189-machine-learning-algorithm-beats-20-lawyers-nda-legal.html> [<https://perma.cc/D5QH-JHWQ>].

<sup>90</sup> See, e.g., Lincoln Mead, *AI Strengthens Your Legal Analytics*, A.B.A., [https://www.americanbar.org/groups/law\\_practice/publications/law\\_practice\\_magazine/2020/jf2020/jf20mead/](https://www.americanbar.org/groups/law_practice/publications/law_practice_magazine/2020/jf2020/jf20mead/) (last visited Sept. 6, 2021) [<https://perma.cc/VN7S-XYDF>] (noting that a “shining example” of an area where AI can supplement lawyers’ work is legal analytics, defined as “the implementation of established data analysis methodologies, using supporting tools, to common data sets within the field of law to improve efficiency, gain insight and realize greater utility from the available data”).

<sup>91</sup> See, e.g., Matthew Hutson, *Artificial Intelligence Prevails at Predicting Supreme Court Decisions*, SCI. (May 2, 2017, 1:45 PM), <https://www.sciencemag.org/news/2017/05/artificial-intelligence-prevails-predicting-supreme-court-decisions> [<https://perma.cc/B6A3-E8D7>] (“A new study shows that computers can do a better job than legal scholars at predicting Supreme Court decisions, even with less information.”).

<sup>92</sup> Goodman, *supra* note 89, at 154.

sentence.<sup>93</sup> It is fair to say that AI has already thoroughly suffused the practice of law.

In examining some of the potential opportunities and perils of using GPT-3 in the legal field, it is helpful to look at an area of law practice where AI algorithms already dominate. To the surprise, perhaps, of many practitioners who use them daily, AI algorithms fuel the legal research databases that lawyers rely on, databases such as Westlaw and LexisNexis.<sup>94</sup> Most lawyers are not familiar with even the basics of the algorithms that produce the search results they get.<sup>95</sup> As is discussed below, such a lack of understanding hinders their research productivity and may even run afoul of the rules of professional conduct.<sup>96</sup> These algorithms, like all algorithms, reflect a multitude of decisions made by the data scientists and software engineers who created them.<sup>97</sup>

Those decisions have consequences for the end users of the algorithms. Imagine this (highly plausible) scenario: a lawyer is using a legal search engine for the purpose of drafting a brief. Wishing to keep costs down for the client, the lawyer decides to exclusively do the research in one commercial research database. The lawyer does not understand that each company utilizes a different algorithm to facilitate the search and has chosen to prioritize different results.<sup>98</sup> Thus, the

---

<sup>93</sup> See generally Cyphert, *supra* note 60, at 338 (“As of 2015, over sixty different risk assessment tools were used in the sentencing context alone, and more were used for bail determinations and by corrections officials.”).

<sup>94</sup> See generally Susan Nevelow Mart, *The Algorithm as a Human Artifact: Implications for Legal [Re]Search*, 109 L. LIBR. J. 387 (2017) (comparing results of the search algorithms used by the search algorithms in Westlaw, Lexis Advance, Fastcase, Google Scholar, Ravel, and Casetext).

<sup>95</sup> See *id.* at 393 (“[I]n the early days of online searching, most users were unaware of the structure underlying the system. This is almost certainly still true.”).

<sup>96</sup> See *id.* at 391 (“[L]earning to navigate black boxes [of search software algorithms] is part of the ethical duty to do competent research: knowing something about why you received the results that you did is a critical skill.”).

<sup>97</sup> *Id.* at 388 (noting that the choices that “human creators” made about “how the algorithm would work [] have implications for the search results returned to the researcher”).

<sup>98</sup> *Id.* at 388-89 (noting that a variety of choices are made in how the search algorithm will operate, and noting that “[i]f the search entered into a legal database has five terms, and only four terms appear, how will the algorithm treat the search? If the algorithm is strict, it will return only results with exactly those five terms. But the algorithm can be adjusted so that results with four of the terms will appear in the results set. The algorithm is set to determine how close those words have to be to each other to be returned in the top results. The programming team decides which of the search terms entered are automatically stemmed and which are not. Only the team knows which legal phrases are recognized by the algorithm without quotation marks around

lawyer does not understand that various databases would yield different results even when given the exact same inputs.<sup>99</sup> Since this lawyer did not understand the technology being used, the lawyer made a choice that might initially save money for the client but at the cost of weakening the brief (and potentially costing the client money in the long term). This cautionary tale is worth bearing in mind, as the implications of misusing GPT-3 could be even more severe.

*B. Why GPT-3 May Be Especially Important for Legal Practice*

GPT-3 is now available for commercial use, and OpenAI reported in a March 2021 blogpost that the technology powers more than 300 apps and generates an average of 4.5 billion words per day.<sup>100</sup> Although, as noted above, AI is hardly new to the practice of law, two features of GPT-3's design make it especially likely to be adopted for use in the legal sector. First, it is a "creation engine" that actually generates text, rather than one that simply sorts or classifies data.<sup>101</sup> Because one of the most important "products" lawyers produce is writing (contacts, motions, etc.), GPT-3 will be especially useful to them.<sup>102</sup>

Second, as noted above, GPT-3 is already pretrained on a massive dataset, and it also has the ability for users to train it on their own

---

the phrase and how many preexisting legal phrases are added to the search without user input").

<sup>99</sup> See *id.* at 390 (concluding that when researchers studied six different legal databases — Casetext, Fastcase, Google Scholar, Lexis Advance, Ravel, and Westlaw — by entering the same search terms into each, there was "hardly any overlap in the cases that appear in the top ten results returned by each database," despite the fact that the researchers used "jurisdictional limits [to create] a unique opportunity to compare how different algorithms process the same search in the same set of documents").

<sup>100</sup> OpenAI & Ashley Pilipiszyn, *GPT-3 Powers the Next Generation of Apps*, OPENAI BLOG (Mar. 25, 2021), <https://openai.com/blog/gpt-3-apps/> [<https://perma.cc/4W6V-Q3NL>].

<sup>101</sup> Rudy DeFelice, *A New AI Model Focused on Doing, Not Thinking — and That's Great News for Lawyers*, LAW.COM: TEX. LAW. (Sept. 8, 2020, 6:42 PM), <https://www.law.com/texaslawyer/2020/09/08/a-new-ai-model-focused-on-doing-not-thinking-and-thats-great-news-for-lawyers/> [<https://perma.cc/Y5LM-HA58>] ("AI tools are generally used in an enterprise to find or categorize information. GPT-3 actually creates things and generates the kind of end products typically created by knowledge workers.").

<sup>102</sup> *Id.* ("Legal organizations generate documents as end products. Essentially, documents are our widget.") Mr. DeFelice notes the variety of legal projects GPT-3 could theoretically assist with: "One can imagine GPT-3 being part of the process that creates initial drafts of legal memoranda, contracts, policy manuals, HR documents, RFP's and audit responses, among other things commonly created by finding and patching together prior versions of these documents by people." *Id.*

---

---

dataset as well.<sup>103</sup> GPT-3's creators have called this an "emergent quality," because GPT-3 is capable of recognizing patterns in the inputs it is given and predicting what will follow.<sup>104</sup> Although previous versions of the model could also be tailored to specific tasks, GPT-3 can do this with remarkable ease. Because it is a few-shot learning model, you can personalize GPT-3 much more easily than you could earlier versions.<sup>105</sup> So, lawyers could relatively easily prime GPT-3 for the kind of writing they want it to produce, and they need not be technology experts in order to do that.<sup>106</sup>

One of the beta testers shared a fascinating demonstration where they fed GPT-3 only two prompts that demonstrated "translating" from "Legalese" to "plain English."<sup>107</sup> For example, the first prompt had the following as its "Legal Clause":

The Company and the Founders will provide the Investors with customary representations and warranties examples of which are set out in Appendix 4 and the Founders will provide the Investors with customary non-competition, non-solicitation and confidentiality undertakings.<sup>108</sup>

And the following as its "plain English" translation:

The startup and its founders will provide the usual assurances and guarantees on facts about the business. The founders will also agree not to work for competitors, poach employees or customers when they leave the startup, and respect confidentiality.<sup>109</sup>

---

<sup>103</sup> See Metz, *supra* note 10 ("Before asking GPT-3 to generate new text, you can focus it on particular patterns it may have learned, priming the system for certain tasks.").

<sup>104</sup> *Id.* (quoting Dario Amodei, Vice President for Research at OpenAI).

<sup>105</sup> *Id.* (noting that you can personalize GPT-3 "using just a few examples, as opposed to the thousands of examples and several hours of additional training required by its predecessors").

<sup>106</sup> *Id.* (quoting Ilya Sutskever, OpenAI's chief scientist, as saying "[a]ny layperson can take this model and provide these examples in about five minutes and get useful behavior out of it").

<sup>107</sup> Tefula, *supra* note 86.

<sup>108</sup> *Id.*

<sup>109</sup> *Id.*

Based solely on that prompt and single other similar one, GPT-3 was able to roughly<sup>110</sup> “translate” phrases from Legalese to plain English, and vice versa. For example, when fed the prompt:

Sale of all or substantially all of the assets of the Company or a sale of shares involving a change in control (each, a “Corporate Transaction”) will be treated in the same way as a liquidation and the proceeds of sale will be distributed as set out in paragraph 3. If the holders of Series A Shares have received any Special Dividend it shall be set off against their Liquidation Preference.<sup>111</sup>

GPT-3 “translated” that language into plain English:

If the company is sold, or a new owner takes control, the proceeds of the sale will be distributed as in the liquidation clause above. Any special dividend paid will be treated as an initial payment towards the Series A investors.<sup>112</sup>

### C. *Increasing Access to Justice?*

It is easy to imagine why law firms may be interested in having a tool that takes a first pass at turning legal documents into something more easily understood by a layperson. But there is a hope that tools like GPT-3 could someday be more than efficiency enhancers for law firms and could instead be used to help address the widening gulf between those with the means to hire attorneys and those without. This access to justice problem — those who need legal services but cannot afford them — is well documented. The Legal Services Corporation<sup>113</sup> reports that in 2017, “86% of the civil legal problems reported by low-income Americans in the past year received inadequate or no legal help.”<sup>114</sup> This

---

<sup>110</sup> As the example here makes clear, the “translation” was not always an ideal one. It might still offer an attorney a decent “first draft” at explaining a complicated legal term to a layperson. This could be helpful in a variety of contexts, including in preparing opening or closing statements before a jury.

<sup>111</sup> Tefula, *supra* note 86.

<sup>112</sup> *Id.*

<sup>113</sup> The Legal Services Corporation, or LSC, is “the single largest funder of civil legal aid for low-income Americans in the nation.” LEGAL SERVS. CORP., <https://www.lsc.gov/about-lsc/who-we-are> (last visited July 13, 2021) [<https://perma.cc/GF5Y-JHX4>].

<sup>114</sup> LEGAL SERVS. CORP., EXECUTIVE SUMMARY: THE JUSTICE GAP: MEASURING THE UNMET CIVIL LEGAL NEEDS OF LOW-INCOME AMERICANS 1 (2017), <https://www.lsc.gov/sites/default/files/images/TheJusticeGap-ExecutiveSummary.pdf> [<https://perma.cc/C6QQ-X4VN>].

access to justice gulf was exacerbated by the economic recession of 2008,<sup>115</sup> further worsened by the COVID-19 pandemic,<sup>116</sup> and is worse in the United States than in other countries.<sup>117</sup>

Might GPT-3 be able to help attorneys provide effective legal services to a larger group of low-income people? A lot of excitement already surrounds the idea of artificial intelligence helping to address the justice gap; indeed the Legal Services Center held a summit on this very topic in 2013.<sup>118</sup> Scholars have acknowledged that artificial intelligence will not fully solve the justice gap, but have nonetheless predicted it could make a real difference.<sup>119</sup> Imagine a chatbot<sup>120</sup> powered by GPT-3 that people seeking legal information could access to ask questions about

---

<sup>115</sup> See, e.g., Raymond H. Brescia, Walter McCarthy, Ashley McDonald, Kellan Potts & Cassandra Rivais, *Embracing Disruption: How Technological Change in the Delivery of Legal Services Can Improve Access to Justice*, 78 ALB. L. REV. 553, 588 (2015) (“The ‘Great Recession’ of 2008 increased the need for legal services for low- and moderate-income individuals.”).

<sup>116</sup> See, e.g., WORLD JUST. PROJECT, THE COVID-19 PANDEMIC AND THE GLOBAL JUSTICE GAP 3 (2020), <https://worldjusticeproject.org/sites/default/files/documents/WJP-COVID%2BGlobal-Justice-Gap-final-10.15.pdf> [https://perma.cc/CL4H-VND9] (“The pandemic is significantly worsening an already serious global gap in access to justice.”).

<sup>117</sup> Deborah L. Rhode, *Access to Justice: An Agenda for Legal Education and Research*, 62 J. LEGAL EDUC. 531, 534 (2013) (“About a quarter of middle-income individuals and between a fifth to half of low-income individuals [took no action on their legal problems] in the United States, compared with 5 percent to 18 percent in most other countries.”).

<sup>118</sup> See LEGAL SERVS. CORP., REPORT OF THE SUMMIT ON THE USE OF TECHNOLOGY TO EXPAND ACCESS TO JUSTICE 1 (2013), [https://www.lsc.gov/sites/default/files/LSC\\_Tech%20Summit%20Report\\_2013.pdf](https://www.lsc.gov/sites/default/files/LSC_Tech%20Summit%20Report_2013.pdf) [https://perma.cc/N792-GGLN] (“The Legal Services Corporation (LSC) has found through its experience with its Technology Initiative Grant program that technology can be a powerful tool in narrowing the justice gap — the difference between the unmet need for civil legal services and the resources available to meet that need.”).

<sup>119</sup> See, e.g., Brescia et al., *supra* note 115, at 592 (“[I]nvesting time, money, and research into new and innovative ways to provide legal aid and representation to low- and middle-income individuals can help bridge the justice gap.”); Anjanette H. Raymond & Scott J. Shackelford, *Technology, Ethics, and Access to Justice: Should an Algorithm Be Deciding Your Case?*, 35 MICH. J. INT’L L. 485, 492 (2014) (concluding that a well-designed online dispute resolution system “can increase individuals’ access to justice”); Drew Simshaw, *Ethical Issues in Robo-Lawyering: The Need for Guidance on Developing and Using Artificial Intelligence in the Practice of Law*, 70 HASTINGS L.J. 173, 180 (2018) (“AI will be an even more impactful force [on the justice gap] than previous tools, and has the potential to magnify and transform benefits of existing technologies.”).

<sup>120</sup> “A chatbot is a virtual software program in which the user communicates with a virtual machine that imitates human conversations through voice and/or text.” Sherley E. Cruz, *Coding for Cultural Competency: Expanding Access to Justice with Technology*, 86 TENN. L. REV. 347, 364 (2019).

their potential claims. Those seeking assistance could ask questions about whether the actions of a landlord are legal in their state and how to file a claim if they are not. GPT-3's demonstrated ability to help "translate" from legalese to plain English could be especially helpful here. For example, the chatbot could help recognize that plain English terms like "kicked out" mean the same thing as Legalese terms like "eviction" and respond accordingly, directing the user to the best guidance and making appropriate referrals to legal services in the area.

Chatbots like this are already being studied as a possible way to increase access to justice.<sup>121</sup> "Learned Hands," a machine learning labeling game, is a joint project between the Stanford Legal Design Lab and Suffolk Law School's Legal Innovation and Technology Lab. Lawyers and law students can "play" a game where they help train an algorithm to recognize legal issues in people's stories,<sup>122</sup> a sort of law school issue spotter exam for the real world. The ultimate goal is to help "make a Rosetta Stone for legal help — linking the legal help guides that courts and legal aid groups offer to the people who are searching for help."<sup>123</sup> However, there is good reason to proceed with great caution before using GPT-3 for legal chatbots.

### III. SPECIFIC ETHICAL CONSIDERATIONS FOR LAWYERS USING GPT-3

As lawyers provide legal services to their clients, they are governed by a variety of ethical rules: rules of local or specialized practice bar associations, the individual practices and rules of judges, and rules of specific courts. This Article focuses on the Model Rules of Professional Conduct ("Model Rules"), which were first promulgated by the American Bar Association in 1983.<sup>124</sup> Since then, they have been adopted in some form by nearly every state in the United States<sup>125</sup> and

---

<sup>121</sup> See *id.* ("Chatbots expand access to justice by providing self-represented litigants with 'personalized' legal guidance to help identify legal issues.").

<sup>122</sup> LEARNED HANDS, <https://learnedhands.law.stanford.edu/> (last visited July 13, 2021) [<https://perma.cc/E4VL-JB7M>] ("The labeled datasets and machine learning models you help us create will be used to improve how courts, legal aid groups, and others can serve people online, when they're looking for help.").

<sup>123</sup> *Id.*; see also Cruz, *supra* note 120, at 364 ("Chatbots can also connect individuals to legal service providers after the program helps the individual identify their legal issue.").

<sup>124</sup> See David Hricik & Jae Ellis, *Disparities in Legal Ethical Standards Between State and Federal Judicial Systems: An Analysis and a Critique*, 13 GEO. J. LEGAL ETHICS 577, 580-81 (2000) (tracing the history of Model Rules).

<sup>125</sup> For a list of when states adopted the Model Rules, see *Alphabetical List of Jurisdictions Adopting Model Rules*, A.B.A., [https://www.americanbar.org/groups/professional\\_responsibility/publications/model\\_rules\\_of\\_professional\\_conduct/alpha](https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/alpha)



have occasionally been amended as well. The Model Rules cover a wide array of lawyer behavior and impose specific duties and obligations.

As the amendments in 2012 make clear, the Model Rules *do* specifically address lawyers' duties and obligations with respect to technology, including artificial intelligence.<sup>126</sup> The ABA House of Delegates reaffirmed in 2019 that lawyers need to take their ethical considerations regarding the use of AI seriously, passing Resolution 112:

RESOLVED, That the American Bar Association urges courts and lawyers to address the emerging ethical and legal issues related to the usage of artificial intelligence (“AI”) in the practice of law including: (1) bias, explainability, and transparency of automated decisions made by AI; (2) ethical and beneficial usage of AI; and (3) controls and oversight of AI and the vendors that provide AI.<sup>127</sup>

Despite the Model Rules' 2012 amendments and the 2019 resolution, scholars still argue that the Model Rules should be further updated to clarify the impact on technology and especially the use of artificial intelligence in the practice of law.<sup>128</sup> This Article attempts to provide a starting place for lawyers who are thinking through their ethical obligations when using GPT-3 in their practice of law. The current Model Rules already impose certain duties on attorneys who wish to use AI like GPT-3.

Three model rules in particular — Rule 1.1 regarding competence, Rule 5.3 regarding supervising nonlawyer assistance, and Rule 8.4 regarding bias — provide certain guidance (and potentially raise certain issues) for lawyers who wish to utilize GPT-3 in their practice. Although the focus of this Article is on GPT-3, many of the principles are

---

[\\_list\\_state\\_adopting\\_model\\_rules/](#) (last updated Mar. 28, 2018) [<https://perma.cc/7N7S-U3YJ>].

<sup>126</sup> See, e.g., Anita Bernstein, *Minding the Gaps in Lawyers' Rules of Professional Conduct*, 72 OKLA. L. REV. 125, 130 (2019) (noting that prior to 2012, the “ABA view of competence had been starkly silent on technology,” but that the Rules now specifically address technology); see also *id.* at 130-44 (discussing changes made in 2012 to Rules 1.1 and 5.3).

<sup>127</sup> AM. BAR ASS'N HOUSE OF DELEGATES, RESOLUTION 112 (2019), <https://www.americanbar.org/content/dam/aba/directories/policy/annual-2019/112-annual-2019.pdf> [<https://perma.cc/2LZ3-JSES>].

<sup>128</sup> See, e.g., Katherine Medianik, *Artificially Intelligent Lawyers: Updating the Model Rules of Professional Conduct in Accordance with the New Technological Era*, 39 CARDOZO L. REV. 1497, 1501-02 (2018) (arguing that the “existing language and content of the Model Rules is outdated and does not account for technological advancement,” and proposing amendments to the Model Rules to “to guide lawyers in situations where they interact with AI tools”).

applicable to other AI models (and indeed to GPT-4, GPT-5, etc., should they be developed). Integration of artificial intelligence into the legal field creates many interesting questions beyond the three rules on which this Article focuses. For example, could reliance on artificial intelligence tools run afoul of Model Rule 2.1, which requires that lawyer exercise independent judgment in the practice of law?<sup>129</sup> Or is there actually an affirmative duty under Model Rule 1.5 that lawyers use AI like GPT-3 when doing so could save clients' money?<sup>130</sup> At least one judge in Ontario has ruled there might be.<sup>131</sup> Both are intriguing questions that are outside the scope of this Article.

---

<sup>129</sup> MODEL RULES OF PRO. CONDUCT r. 2.1 (AM. BAR ASS'N 2020) ("In representing a client, a lawyer shall exercise independent professional judgment and render candid advice." Such advice may be about law or "other considerations such as moral, economic, social and political factors, that may be relevant to the client's situation."); see also Simshaw, *supra* note 119, at 204 ("On a more abstract level, as lawyers become increasingly reliant on intelligent systems, it draws into question the extent to which their professional judgment is 'independent.' This is especially true if they do not fully understand and were not involved with the design of the system, and therefore cannot make independent judgments based on the AI's output.").

<sup>130</sup> Model Rule of Professional Conduct 1.5 requires that lawyers not charge unreasonable fees, and notes that factors for determining the reasonableness of fees include the time and labor required. Scholars have suggested that if AI can save significant billable hours, there is an arguable duty for lawyers to use it. See, e.g., Roy D. Simon, *Artificial Intelligence, Real Ethics*, 90 N.Y. STATE BAR J. 34, 37 (2018) ("I think you are not charging an excessive fee if you continue using your customary methods instead of using a new-fangled AI product, but soon most lawyers will be using AI products and services for certain types of work (such as the cite-checking products discussed earlier), and charging for 10 hours of your time to do work that AI could do in 10 minutes sounds like an excessive fee to me. You have to keep abreast of the benefits of technology that applies to your practice. . . . Do you have a duty to alert your clients to the option of using AI products that may save substantial fees or arrive at quicker or more accurate results? Right now the answer to that question is unclear — but before long, practicing law without using AI will be like practicing law with an Underwood manual typewriter, and you will have to tell your clients that there is a better, cheaper, faster way."); see also Ed Walters, *The Model Rules of Autonomous Conduct: Ethical Responsibilities of Lawyers and Artificial Intelligence*, 35 GA. STATE U. L. REV. 1073, 1076 (2019) ("[A]s the quality of work product created by lawyers augmented with AI surpasses the work created without AI, it is clear that lawyers will soon have a professional responsibility to employ new techniques.").

<sup>131</sup> One court in Ontario slashed a party's costs request because the lawyer had not used AI when conducting legal research. The judge ruled that "[t]here was no need for outsider or third party research. If artificial intelligence sources were employed, no doubt counsel's preparation time would have been significantly reduced." *Cass v. 1410088 Ontario Inc.*, 2018 O.N.S.C. 6959, para. 34 (Can. Ont. Sup. Ct. J.).

A. *Rule 1.1 and the General Duty of Competence*

1. Competence Defined for AI

Lawyers must competently practice law. Specifically, Model Rule 1.1 requires that lawyers “shall provide competent representation to a client. Competent representation requires the legal knowledge, skill, thoroughness and preparation reasonably necessary for the representation.”<sup>132</sup> In 2012, the drafters of the Model Rules clarified in Comment 8 to Rule 1.1 that this duty of competence includes a duty to remain abreast of changes in the practice of law, “including the benefits and risks associated with relevant technology.”<sup>133</sup> In describing the amendment to Comment 8, one reporter succinctly noted that it meant that “lawyers can’t be Luddites.”<sup>134</sup> As Comment 8 suggests, the 2012 amendments were specifically made, in part, to address the growing use of technology in the practice of law.<sup>135</sup> This new emphasis on technical competence as an ethical duty for lawyers “propelled [lawyers] headlong into a complex world of fast-changing technological growth.”<sup>136</sup>

In adopting various versions of Rule 1.1, some states have gone beyond what the Model Rule requires and are more specific and prescriptive. For example, Florida says that lawyers must engage in continuing education about technology in order to competently practice law.<sup>137</sup> In adopting the new Comment 8, West Virginia changed the Model Rule language that “a lawyer *should* keep abreast” to “a lawyer

---

<sup>132</sup> MODEL RULES OF PRO. CONDUCT r. 1.1 (AM. BAR ASS’N 2020).

<sup>133</sup> MODEL RULES OF PRO. CONDUCT r. 1.1 cmt. 8.

<sup>134</sup> Debra Cassens Weiss, *Lawyers Have Duty to Stay Current on Technology’s Risks and Benefits*, *New Model Ethics Comment Says*, A.B.A. J. (Aug. 6, 2012, 7:46 PM CDT), [https://www.abajournal.com/news/article/lawyers\\_have\\_duty\\_to\\_stay\\_current\\_on\\_technologys\\_risks\\_and\\_benefits](https://www.abajournal.com/news/article/lawyers_have_duty_to_stay_current_on_technologys_risks_and_benefits) [https://perma.cc/SW36-Y644].

<sup>135</sup> See Simshaw, *supra* note 119, at 196 (noting that the 2012 amendments updated “the black letter and commentary of several key model rules in order to take into account the increased role of technology in the profession”).

<sup>136</sup> Jason Tashea & Nicholas Economou, *Be Competent in AI Before Adopting, Integrating It into Your Practice*, A.B.A. J. (Apr. 23, 2019, 6:30 AM CDT), <http://www.abajournal.com/lawscribbler/article/before-lawyers-can-ethically-adopt-and-integrate-ai-into-their-practices-they-must-first-be-competent> [https://perma.cc/45P6-B72G].

<sup>137</sup> See Medianik, *supra* note 128, at 1515 (citing Fla. Bar Pro. Ethics Comm., Op. 06-2 (2006)) (discussing duties with respect to metadata in emails and noting that lawyers’ professional obligations “may necessitate a lawyer’s continuing training and education in the use of technology”).

must keep abreast” of changes to the practice of law, including “the benefits and risks associated with relevant technology.”<sup>138</sup>

Other states have tried to clarify what competence means through ethics opinions. For example, an Arizona Ethics Opinion reminded lawyers that their duty of competence meant that if they lacked “the training or experience required to act competently with regard to computer security,” that such competence was nonetheless readily available.<sup>139</sup> Attorneys are encouraged to do their own research to learn more about computer security and/or to work with experts to improve their security practices.<sup>140</sup>

California also urges lawyers who lack certain technological competence to enlist the help of experts as one of several options lawyers have. In a formal opinion from its professional responsibility committee, the California Bar provides that “[a]n attorney lacking the required competence for e-discovery issues has three options: (1) acquire sufficient learning and skill before performance is required; (2) associate with or consult technical consultants or competent counsel; or (3) decline the client representation.”<sup>141</sup> Of course, as will be explored further below, attorneys who lack certain technological competences and so associate with experts still retain the duty to supervise that expert’s work. Consultation with such an expert “does not absolve an attorney’s obligation to supervise the work of the expert under [California’s duty to supervise], which is a non-delegable duty

---

<sup>138</sup> Jamie J. Baker, *Beyond the Information Age: The Duty of Technology Competence in the Algorithmic Society*, 69 S.C. L. REV. 557, 563 (2018) (“By purposefully changing the language from ‘should’ to ‘must,’ West Virginia has signaled a stronger ethical duty to its lawyers.”). Compare MODEL RULES OF PRO. CONDUCT r. 1.1 cmt. 8 (“[A] lawyer should keep abreast” of “the benefits and risks associated with relevant technology . . .”), with W. VA. RULES OF PRO. CONDUCT r. 1.1 cmt. 8 (2015) (“[A] lawyer must keep abreast of the benefits and risks associated with relevant technology . . .”).

<sup>139</sup> State Bar of Ariz. Comm. on the Rules of Pro. Conduct, Formal Op. 05-04 (2005), <https://tools.azbar.org/RulesofProfessionalConduct/ViewEthicsOpinion.aspx?id=523> [<https://perma.cc/XW5A-VB8Z>].

<sup>140</sup> *Id.* (“Much information can be obtained through the internet by an attorney with sufficient time and energy to research and understand these systems. Alternatively, experts are readily available to assist an attorney in setting up the firm’s computer systems to protect against theft of information and inadvertent disclosure of client confidences.”)

<sup>141</sup> State Bar of Cal. Standing Comm. on Pro. Resp. & Conduct, Formal Op. No. 2015-193, at 1 (2015), [http://www.calbar.ca.gov/Portals/0/documents/ethics/Opinions/CAL%202015-193%20%5B11-0004%5D%20\(06-30-15\)%20-%20FINAL1.pdf](http://www.calbar.ca.gov/Portals/0/documents/ethics/Opinions/CAL%202015-193%20%5B11-0004%5D%20(06-30-15)%20-%20FINAL1.pdf) [<https://perma.cc/C6KT-C486>].

belonging to the attorney who is counsel in the litigation, and who remains the one primarily answerable to the court.”<sup>142</sup>

As is clear from the text of Rule 1.1, Comment 8, and the various state ethics opinions, an attorney cannot use technology that they do not understand. In teasing out what this duty of competence means with respect to the use of artificial intelligence in the practice of law, scholars have focused largely on two ethical obligations: first, lawyers must have at least a rudimentary understanding of the technology (which can involve hiring a technical expert to help them learn more and vet products); and second, they must not blindly adopt an AI program’s outputs without some level of supervision and/or skepticism.<sup>143</sup> The challenges that rise with “supervising” AI are addressed below in the discussion around Rule 5.3. Regarding understanding the technology, most scholars advocate only for a “basic understanding” of the technology<sup>144</sup> (though some argue that lawyers should be required to attend mandatory CLEs focused on legal technology,<sup>145</sup> a recommendation this Article echoes). For many attorneys, even the low bar of a basic understanding of AI programs is going to be difficult to clear. AI programs, especially those fueled by machine learning, can be quite opaque and difficult for even a technical expert to understand and explain.<sup>146</sup> And, of course, most lawyers are not technical experts.<sup>147</sup>

---

<sup>142</sup> *Id.* at 5.

<sup>143</sup> See Simon, *supra* note 130, at 35 (suggesting that lawyers “(1) hire an expert to vet the AI product; (2) learn what the AI product can (and can’t) do; and (3) double-check the output of the AI product”); see also Nicole Yamane, *Artificial Intelligence in the Legal Field and the Indispensable Human Element Legal Ethics Demands*, 33 *GEO. J. LEGAL ETHICS* 877, 883-84 (2020) (noting that lawyers must have a basic understanding of the AI tools they use and must exercise care when using them, which means reviewing the program’s results).

<sup>144</sup> See, e.g., Yamane, *supra* note 143, at 883 (“[L]awyers must have a basic understanding of the AI programs they choose to utilize in their practice.”).

<sup>145</sup> See, e.g., Medianik, *supra* note 128, at 1526 (advocating for an amendment to Model Rule 1.1 to mandate CLEs focused on legal technology and urging the ABA to “establish a ‘Legal Technology’ section as an additional topic of discussion for CLE credits”).

<sup>146</sup> See Will Knight, *The Dark Secret at the Heart of AI*, *MIT TECH. REV.* (Apr. 11, 2017), <https://www.technologyreview.com/2017/04/11/51113/the-dark-secret-at-the-heart-of-ai/> [<https://perma.cc/R8VF-MBZC>].

<sup>147</sup> See Yamane, *supra* note 143, at 883 (“Because AI is a branch of computer science and often involves technical knowledge outside of most lawyers’ expertise, understanding how AI programs operate may be difficult for lawyers.”); Tashea & Economou, *supra* note 136 (lamenting that statistics courses are not a mainstay in most law school curricula and noting that “[t]he science underpinning effective and measurable results of AI is not for the faint of heart. Governed by computer science and

Regardless, lawyers must still “maintain a baseline of knowledge about the AI programs they use,”<sup>148</sup> and this knowledge should include how the AI program reaches its outputs and any limitations of the program.<sup>149</sup>

## 2. Competence and GPT-3

What might such a baseline of knowledge look like for a lawyer who utilizes GPT-3? It will be critical for that lawyer to understand three things: (1) GPT-3 will sometimes produce results that mirror larger racial or other biases in our society, (2) GPT-3 makes silly mistakes, and (3) GPT-3 will tend to uphold the legal status quo and may not therefore be an ideal tool for advocating for change of existing precedent.

First, the lawyer needs to understand the origins of GPT-3’s data training set, and specifically that it was trained in part on internet message boards that often include language that is misogynistic and racist. This massive data training set is, of course, part of why GPT-3 is so effective at predicting text. But any lawyer who utilizes it needs to be aware of the potential for it to produce biased outputs (this is especially important in light of Model Rule 8.4(g)’s prohibition against discrimination and bias in the practice of law, which is covered more fully below).<sup>150</sup>

Second, a lawyer who is utilizing GPT-3 in their practice needs to be aware that the program — despite all of the buzz and despite the fact that it really is a technological advance in the field of generative text — can still make some very silly mistakes. Most seasoned lawyers can share a horror story about a typo in their work, perhaps one caused by autocorrect. For example, one lawyer presented an appellate brief to the Ninth Circuit Court of Appeals, “in which auto-correct had changed

---

statistics, these are complex academic disciplines in which lawyers are generally untrained and cannot become experts on the fly.”).

<sup>148</sup> Yamane, *supra* note 143, at 884 (“Without this baseline of knowledge, lawyers will be unable to use AI programs with full competence, thereby jeopardizing their ability to provide competent representation to their clients.”).

<sup>149</sup> *See id.* (indicating that lawyers need to understand “(1) why the AI program produces its results and (2) what the AI program is and is not capable of”); *see also* Stuart Teicher, *Tech Tock, Tech Tock: The Countdown to Your Ethical Demise*, 31 J. AM. ACAD. MATRIM. LAWS. 481, 498 (2019) (“[I]t appears a lawyer’s duty of competence probably already includes big data. The idea that entities are collecting, sharing, and analyzing data about lawyers and their clients is common knowledge. Being able to understand how that whole process works, at least at a basic level, appears to be necessary to establish minimum levels of competence.”).

<sup>150</sup> *See supra* Part I.C.

‘sua sponte’ to ‘sea sponge,’ resulting in the sentence: ‘[I]t is well settled that a trial court must instruct sea sponge on any defense, including a mistake of fact defense.’”<sup>151</sup> These types of errors, while sometimes funny, can irritate courts,<sup>152</sup> impact contracts in ways that lead to extended litigation,<sup>153</sup> and perhaps in extreme cases even lead to outcomes like wrongful convictions.<sup>154</sup> If a lawyer chooses to use GPT-3 to help draft legal pleadings, a certain amount of caveat emptor is necessary. Lawyers’ duties to supervise AI tools is discussed at greater length below, but knowing that GPT-3 is prone to silly mistakes raises the stakes of using it.

Third, it is important to remember that GPT-3 is a prediction tool. Like all predictive tools, it has a bias toward replicating the past, specifically toward replicating its own data training set.<sup>155</sup> This tendency for prediction tools to keep repeating the past can be seen with recidivism prediction tools that rely on AI. Imagine an algorithm that is trained on historical criminal justice data. If that data is biased because it reflects a racially unfair criminal justice system where Black men were disproportionately likely to be arrested, charged with, and convicted of crimes, then any outputs of the recidivism prediction algorithm will be similarly biased and will predict it is more likely that Black men will reoffend.<sup>156</sup> The tool can only predict that which it has been trained to

---

<sup>151</sup> Robert D. Lang, *From “Sua Sponte” to “Sea Sponge” The Mixed Blessings of Auto-Correct*, N.Y. STATE BAR ASS’N J., July/Aug. 2015, at 28, 29.

<sup>152</sup> See, e.g., Adam Liptak, *Judge Finds a Typo-Prone Lawyer Guilty of Bad Writing*, N.Y. TIMES (Mar. 4, 2004), <https://www.nytimes.com/2004/03/04/us/judge-finds-a-typo-prone-lawyer-guilty-of-bad-writing.html> [<https://perma.cc/VC5K-JZ5J>] (reporting on a federal judge who reduced a lawyer’s fee request because his writing included so many typos).

<sup>153</sup> See, e.g., Debra Cassens Weiss, *No Takebacks? Settlement Offer Missing a Zero Can Be Withdrawn, Appeals Court Says*, A.B.A. J. (Aug. 20, 2020, 11:26 AM CDT), <https://www.abajournal.com/news/article/no-takebacks-paralegals-mistakenly-sent-settlement-offer-can-be-rescinded-appeals-court-says> [<https://perma.cc/6ECX-MHUJ>] (discussing a Florida appeals court ruling regarding a settlement offer that a paralegal mistakenly drafted as \$10,000, rather than \$100,000).

<sup>154</sup> See, e.g., Jennifer Groscup, *Did a Typo Cause a Conviction?*, MONITOR ON PSYCH. (Am. Psych. Assoc.), June 2008, at 20, <https://www.apa.org/monitor/2008/06/jn> [<https://perma.cc/SS3X-VS3P>] (discussing case where use of “and” instead of “or” in jury instructions regarding felony murder charge may have impacted the outcome). Ultimately, on remand from the Supreme Court, the Ninth Circuit denied the habeas petition. *Pulido v. Chrones*, 629 F.3d 1007, 1020 (9th Cir. 2010).

<sup>155</sup> See Mayson, *supra* note 59, at 2251 (“[P]rediction functions like a mirror. The premise of prediction is that, absent intervention, history will repeat itself.”).

<sup>156</sup> *Id.* (“Any form of prediction that relies on data about the past will produce racial disparity if the past data shows the event that we aspire to predict — the target variable — occurring with unequal frequency across racial groups.”).

recognize. Similarly, GPT-3 will tend to predict legal language it has already been trained to recognize. Theoretically, it will suggest that a lawyer include in a motion the same arguments that have previously been included in similar motions. Of course, if the lawyer is working on a routine motion, this may be helpful. But if the lawyer is attempting to chart a creative and novel legal argument to overturn existing precedent, GPT-3's value will likely be limited.

### B. *Supervisory Duties Under Rule 5.3*

#### 1. Supervision Defined for AI

Lawyers have long had a duty to supervise the many “non-lawyers” (paralegals, legal secretaries, accountants, etc.) they work with.<sup>157</sup> This duty is codified in the Model Rule 5.3, which was initially entitled “Responsibilities Regarding Non-Lawyer Assistants.” When Model Rule 5.3 was first promulgated by the ABA in 1983, Microsoft had just introduced its new software “Word”<sup>158</sup> and just 10% of adults said they had a home computer.<sup>159</sup> At the time of the Rule’s passage, therefore, it is unlikely that the drafters were contemplating the best way for a lawyer to “supervise” a non-lawyer AI system capable of drafting volumes of legal documents like GPT-3. However, by the time Model Rule 5.3 was amended in 2012, it was renamed (changing the title from “Responsibilities Regarding Non-Lawyer Assistants” to “Responsibilities Regarding Non-Lawyer Assistance”) to make clear that lawyers *do* indeed have a duty to supervise non-human AI if they utilize it.<sup>160</sup>

The Rule in various portions refers to a “person” when referring to nonlawyer assistance. For example, in subpart (b), which provides that

---

<sup>157</sup> See Douglas R. Richmond, *Watching Over, Watching Out: Lawyers’ Responsibilities for Nonlawyer Assistants*, 61 U. KAN. L. REV. 441, 441 (2012) (listing at length the various nonlawyer assistants lawyers regularly work with). See generally MODEL RULES OF PRO. CONDUCT r. 5.3 (AM. BAR ASS’N 2020) (detailing responsibilities regarding nonlawyer assistance).

<sup>158</sup> *Timeline of Computer History: 1983*, COMPUT. HIST. MUSEUM, <https://www.computerhistory.org/timeline/1983/> (last visited July 13, 2021) [<https://perma.cc/NPT3-EVTS>].

<sup>159</sup> Susannah Fox & Lee Rainie, *Part 1: How the Internet Has Woven Itself into American Life*, PEW RSCH. CTR. (Feb. 27, 2014), <https://www.pewresearch.org/internet/2014/02/27/part-1-how-the-internet-has-woven-itself-into-american-life/> [<https://perma.cc/Q66T-U7SY>].

<sup>160</sup> See AM. BAR ASS’N HOUSE OF DELEGATES, *supra* note 127, at 6 (“In 2012, the title of Model Rule 5.3 was changed from ‘Responsibilities Regarding Nonlawyer Assistants’ to ‘Responsibilities Regarding Nonlawyer Assistance.’ The change clarified that the scope of Rule 5.3 encompasses nonlawyers whether human or not.”).



---

“a lawyer having direct supervisory authority over the nonlawyer shall make reasonable efforts to ensure that the person’s conduct is compatible with the professional obligations of the lawyer. . . .” ABA Resolution 112 nonetheless makes clear that AI is encompassed by the rule, noting that “the scope of Rule 5.3 encompasses nonlawyers whether human or not.”<sup>161</sup> Further, Comment 3 to Rule 5.3 includes as an example of nonlawyer assistance “using an Internet-based service to store client information.”<sup>162</sup> In adopting their own version of Rule 5.3, several states have replaced the word “person” with the word “nonlawyer” to make it clear that AI is encompassed by the Rule.<sup>163</sup>

It is important to note that Rule 5.3 is not a vicarious liability statute: lawyers are not automatically responsible for the actions of the nonlawyers they work with simply by virtue of their relationship to them. Rather, Rule 5.3 imposes on lawyers a specific duty to take *reasonable* efforts to supervise the work of the nonlawyer.<sup>164</sup> Further, the question of what is “reasonable” turns, in part, on “the education, experience and reputation of the nonlawyer,” and “the nature of the services involved . . . .”<sup>165</sup> As part of this reasonable supervision of a nonlawyer, the attorney should “communicate directions appropriate under the circumstances to give reasonable assurance that the nonlawyer’s conduct is compatible with the professional obligations of the lawyer.”<sup>166</sup>

## 2. Supervision and GPT-3

Clearly, lawyers who want to use GPT-3 in their practice will have to supervise it. AI experts warn that GPT-3 should *always* be “babysat” by a human, whether at use in a legal setting or not.<sup>167</sup> But what does such supervision actually look like? At a minimum, GPT-3 cannot be used to produce writing that is presented to clients or courts without a human first reviewing the text to make sure it is accurate and appropriate.

---

<sup>161</sup> *Id.*

<sup>162</sup> MODEL RULES OF PRO. CONDUCT r. 5.3 cmt. 3.

<sup>163</sup> *See* Medianik, *supra* note 128, at 1521.

<sup>164</sup> *See, e.g.*, MODEL RULES OF PRO. CONDUCT r. 5.3 cmt. 3 (“[A] lawyer must make reasonable efforts to ensure that the services are provided in a manner that is compatible with the lawyer’s professional obligations.”).

<sup>165</sup> *Id.*

<sup>166</sup> *Id.*

<sup>167</sup> Thornhill, *supra* note 28 (quoting Shannon Vallor, a professor of the ethics of data and AI at the University of Edinburgh, as saying, “For now, GPT-3 needs a human babysitter at all times to tell it what kinds of things it shouldn’t say.”).

For example, lawyers should not use GPT-3 to power chatbots on their websites to interact with potential clients, a worrisome possibility that is discussed at greater length in Part III.C. A lawyer who signs and submits to a court a document prepared by a paralegal or secretary without closely reading it is held responsible for the contents of that document and can be disciplined for it.<sup>168</sup> Likewise, a lawyer who does not carefully review any writing produced by GPT-3 before passing it along to a court is likewise going to be subject to discipline. A lawyer who fails to supervise an AI tool in accordance with Rule 5.3 is not off the hook merely because the language was not the product of a human.<sup>169</sup>

The consequences for failing to supervise AI like GPT-3 go beyond disciplinary actions that call into question an attorney's competence. Failing to supervise GPT-3 could potentially raise unauthorized practice of law issues. The unsupervised use of GPT-3 by lawyers (or any use of GPT-3 by nonlawyers to perform work considered the practice of law) may be considered to be the unauthorized practice of law in violation of Model Rule 5.5.<sup>170</sup> It is important to note that this area of the law is underdeveloped, and it is difficult to predict with any accuracy what a court or state ethics board<sup>171</sup> might do. "While there have been lawsuits against AI program developers, claiming they engaged in the unauthorized practice of law, legal precedent on this matter is still new and murky."<sup>172</sup>

In a 2015 opinion, the Second Circuit suggested that, at least in the context of document review, any work that can be performed by a

---

<sup>168</sup> For example, Rule 11 of the Federal Rules of Civil Procedure requires that for represented parties, each pleading or motion filed with the court be signed by an attorney of record. FED. R. CIV. P. 11. By signing, the lawyer certifies that the pleadings are not frivolous, have evidentiary support, etc.

<sup>169</sup> See Taylor B. Schaefer, *The Ethical Implications of Artificial Intelligence in the Law*, 55 GONZ. L. REV. 221, 232 (2020) ("[A] court would not likely accept an excuse that e-filing software failed to file an important filing as the attorney has a duty to verify that their work is done competently.").

<sup>170</sup> Model Rule 5.5(a) provides that "[a] lawyer shall not practice law in a jurisdiction in violation of the regulation of the legal profession in that jurisdiction, or assist another in doing so." MODEL RULES OF PRO. CONDUCT r. 5.5 (AM. BAR. ASS'N 2020); see also Schaefer, *supra* note 169, at 233 ("An attorney who has created or participated in the creation and operation of a chatbot or similar service must also be wary of potentially violating Rule 5.5(a).").

<sup>171</sup> As the Second Circuit recognized in the *Lola* decision, "the definition of 'practice of law' is 'primarily a matter of state concern,' . . . [since] '[r]egulating the 'practice of law' is traditionally a state endeavor.'" *Lola v. Skadden, Arps, Slate, Meagher & Flom LLP*, 620 F. App'x 37, 41-42 (2d Cir. 2015) (internal citations omitted).

<sup>172</sup> Yamane, *supra* note 143, at 887.

“machine” is per se *not* the practice of law.<sup>173</sup> The court appeared to give great weight to the fact that the work at issue in *Lola* did not involve the exercise of any independent legal judgment.<sup>174</sup> By application, an attorney who is properly supervising GPT-3, who is using independent legal judgment, would not be participating in the unauthorized practice of law. But any person — attorney or not — who blindly uses the tool to draft legal documents or to provide legal advice may well be.<sup>175</sup> Part IV below explores more deeply how lawyers can be better trained and supported as they attempt to properly supervise AI like GPT-3.

### C. Bias and Rule 8.4

#### 1. Bias Defined for AI

Because discrimination by lawyers “undermine[s] confidence in the legal profession and the legal system,”<sup>176</sup> the Model Rules deem it professional misconduct for lawyers to engage in conduct that is harassment or discriminatory while practicing law.<sup>177</sup> Model Rule 8.4(g) includes many protected classes with respect to discrimination,

---

<sup>173</sup> *Lola*, 620 F. App’x at 45 (“[A]n individual who, in the course of reviewing discovery documents, undertakes tasks that could otherwise be performed entirely by a machine cannot be said to engage in the practice of law.”).

<sup>174</sup> *Id.* at 44-45 (noting that courts in North Carolina, Nevada, Colorado, Oregon, Illinois, and New York have held that the practice of law involves “at least a modicum of independent legal judgment”).

<sup>175</sup> Although a full exploration of the contours of the unauthorized practice of law with respect to AI are outside of the scope of this paper, other scholars have addressed this issue. See, e.g., Simshaw, *supra* note 119, at 178 (“On the legal self-help front, courts, state legislatures, and bar associations in the near term will have to decide whether increasingly sophisticated [AI] services . . . constitute the unauthorized practice of law.”); see also Schaefer, *supra* note 169, at 234 (noting that when tools like chatbots are developed without attorneys, state legislatures are left to define whether that is the unauthorized practice of law, since laypersons are not bound by the Model Rules. “State legislatures often look to state bar associations to define the unauthorized practice of law. In many states, injunctions are becoming more common as a remedy for the unauthorized practice of law.”); Michael Simon, Alvin F. Lindsay, Loly Sosa & Paige Comparato, *Lola v. Skadden and the Automation of the Legal Profession*, 20 YALE J.L. & TECH. 234, 262 (2018) (noting that “a few state bars have tackled the issue [of AI products as unauthorized practice of law], though not conclusively,” and sharing various approaches that states have taken).

<sup>176</sup> MODEL RULES PRO. CONDUCT r. 8.4(g) cmt. 3 (AM. BAR. ASS’N 2020).

<sup>177</sup> MODEL RULES PRO. CONDUCT r. 8.4(g) (“It is professional misconduct for a lawyer to . . . engage in conduct that the lawyer knows or reasonably should know is harassment or discrimination on the basis of race, sex, religion, national origin, ethnicity, disability, age, sexual orientation, gender identity, marital status or socioeconomic status in conduct related to the practice of law.”).

including on the basis of race, sex, gender identity and socioeconomic status.<sup>178</sup> A comment to the Rule defines discrimination as including “harmful verbal or physical conduct that manifests bias or prejudice towards others.”<sup>179</sup> The lawyer need not know that the conduct is discriminatory; it is enough under Rule 8.4(g) if the lawyer *reasonably should have known* that the conduct was discriminatory.<sup>180</sup>

Rule 8.4 is sometimes referred to as a “catch-all” provision,<sup>181</sup> and the discrimination language found in part (g) was only recently added in 2016, after several earlier failed attempts to include the language.<sup>182</sup> As with all of the Model Rules, individual states have varied in their decisions about whether and how to adopt Rule 8.4(g).<sup>183</sup>

## 2. Bias and GPT-3

It is not hard to imagine a scenario involving a lawyer using GPT-3 that would implicate Rule 8.4(g). As noted previously, one of the possible applications of GPT-3 in the legal field is for chatbots that could help direct visitors to a website to the proper legal services for them.<sup>184</sup> This application may be especially intriguing (and problematic) in attempting to address the access to justice gap.

Imagine a legal chatbot powered by GPT-3. Remember that GPT-3 learned to write from, among other sources, the subthreads of Reddit.<sup>185</sup> Now imagine that the chatbot is asked a question about a potential employment discrimination claim, based on race or gender. Remember

---

<sup>178</sup> *See id.*

<sup>179</sup> MODEL RULES PRO. CONDUCT r. 8.4(g) cmt. 3.

<sup>180</sup> MODEL RULES PRO. CONDUCT r. 8.4(g) (providing that it is professional misconduct for a lawyer to engage in conduct that “the lawyer knows or *reasonably should know*” is discriminatory (emphasis added)).

<sup>181</sup> Bernstein, *supra* note 126, at 134.

<sup>182</sup> *See, e.g.,* Veronica Root Martinez, *Combating Silence in the Profession*, 105 VA. L. REV. 805, 811 (2019) (noting that advocates first attempted to formally adopt a rule that prohibits discrimination in the practice of law in 1994, but that that attempt failed, “making the successful passage of Model Rule 8.4(g) in 2016 an apparent victory for those who spent years working to get broad-based support within the bar to address issues of diversity and discrimination”).

<sup>183</sup> *Id.* at 811-12 (noting that “[s]tates have adopted the rule, adopted a less aggressive version of the rule, and formally rejected the rule,” and that some state attorneys general have argued that it is unconstitutional).

<sup>184</sup> *See supra* Part II.C.

<sup>185</sup> Other commentators have speculated that chatbots powered by AI could pose Rule 8.4(g) issues for attorneys. *See* Sharon D. Nelson & John W. Simek, *The ABA Tackles Artificial Intelligence and Ethics*, LAW PRAC., Jan./Feb. 2020, at 26, 27 (“Imagine an AI chatbot on a lawyer’s website [writing racist and sexist text in the same way a Microsoft chatbot did in 2016]. Scary, huh?”).

that the creators of GPT-3 have already acknowledged that it has a tendency to use more sexualized language with respect to women and to use more negative language when writing about Black people. What might GPT-3 say in response to this potential client? The outcome has real potential to be disastrous.

We need not operate entirely in the world of hypothetical, as researchers using GPT-3 during beta testing have already shown that it can prove disastrous as a chatbot. A group of French doctors and machine learning engineers developed a medical chatbot that was powered by GPT-3.<sup>186</sup> The chatbot performed decently well at administrative tasks such as scheduling appointments and determining insurance benefits.<sup>187</sup> However, when attempting to help with mental health questions, the tool went hugely awry. At times, it was merely perplexing — it told one fake patient that recycling their electronics may help them feel happier.<sup>188</sup> But, the chatbot also gave breathtakingly awful advice — it actually told a fake patient who was contemplating suicide that they should, indeed, kill themselves.<sup>189</sup> The French researchers were quick to note that OpenAI has explicitly warned against using GPT-3 in a high-stakes area like medicine,<sup>190</sup> so there is good reason to be cautious of its use for chatbots in law as well.

Once again, the key to using GPT-3 in a way that does not run afoul of Rule 8.4(g) will focus on understanding the technology and its tendencies and supervising it effectively.<sup>191</sup> Unless and until the

---

<sup>186</sup> Anne-Laure Rousseau, Clément Baudelaire & Kevin Riera, *Doctor GPT-3: Hype or Reality?*, NABLA: BLOG (Oct. 26, 2020), <https://www.nabla.com/blog/gpt-3/> [<https://perma.cc/9N2E-6KGQ>].

<sup>187</sup> *Id.* (concluding that “GPT-3 seemed to work for basic admin tasks such as appointment booking, but when digging a bit we found that the model had no clear understanding of time, nor any proper logic,” and also finding that “GPT-3 could help nurses or patients to quickly find a piece of information in a very long document, like finding insurance benefits for specific medical examinations”).

<sup>188</sup> *Id.* (“The model can also shoot unexpected answers where it suggests recycling more to ease stress . . .”).

<sup>189</sup> *Id.* The fake patient typed, “Should I kill myself?” and the GPT-3 chatbot responded with, “I think you should.” *Id.*

<sup>190</sup> *Id.* (“As Open AI itself warns in GPT-3 guidelines, healthcare ‘is in the high stakes category because people rely on accurate medical information for life-or-death decisions, and mistakes here could result in serious harm’. Furthermore, diagnosing medical or psychiatric conditions falls straight in the ‘unsupported use’ of the model.”).

<sup>191</sup> See Caleb Chaplain & Nisha R. Patel, *The Terminator Argument: The Duty of Competence in Using Artificial Intelligence*, 38 AM. BANKR. INST. J. 28, 28-29 (2019) (reviewing Rule 8.4(g) and concluding that “attorneys might have an ethical obligation to understand the data underlying the machine learning to account for [bias from the AI creator or from its data set]”).

developers of GPT-3 can effectively guard against it producing text that is racist or sexist, it should not be used for chatbot features. Further, any attempts to address access to justice issues with AI like GPT-3 need to be carried out with an eye toward the tendency of such tools to produce outputs that are biased against the very people who are often left behind by the justice gap. Scholars have noted that AI tools can do more harm than good when it comes to marginalized communities.<sup>192</sup> If a tool like GPT-3 is to be used successfully in addressing the access to justice gap, it must be reevaluated and updated with an eye toward more culturally competent design.<sup>193</sup> “Unless the designers deliberately consider the issue of biased schemas within their design, AI may promote implicit biases that negatively impact the communities that are in most need of the help.”<sup>194</sup>

But, lawyers will not be able to remove bias from AI systems, no matter how technically competent they become or how rigorously they supervise the systems. Rather, their duty under Rule 8.4(g) will largely be to understand the tendency toward bias so that they can make informed decisions about when the technology is appropriate in the practice of law and when it should be avoided.

#### IV. ARE THE CURRENT MODEL RULES ADEQUATE?

As the preceding section made clear, there are already Model Rules that provide a foundation for defining lawyers’ ethical duties with respect to the use of AI, like GPT-3, in their practices. Are these existing rules “enough” to help effectively guide lawyers’ behavior with respect to the use of GPT-3 and other AI tools in the practice of law? Arguably, the Rules themselves may be adequate, as many commentators and Part III of this Article have located within them certain duties.<sup>195</sup> However, they are currently too lacking in specificity with respect to the use of AI

---

<sup>192</sup> See, e.g., Emily S. Taylor Poppe, *The Future Is Bright Complicated: AI, Apps & Access to Justice*, 72 OKLA. L. REV. 185, 186 (2019) (“I highlight the potential of legal technology to reproduce, rather than ameliorate, existing social inequalities.”); see also Cruz, *supra* note 120, at 369-70 (noting that, in the criminal justice context, “implicit biases in the AI formulas are skewing the results in ways that negatively impact defendants of color. While not strictly an access to justice issue, the biased results highlight the dangers of using technology that does not account for diversity and cultural associations.”).

<sup>193</sup> See Cruz, *supra* note 120, at 351 (examining “the intersectionality of cross-cultural competence theory and access to justice theory to demonstrate that successful use of legal technology inextricably requires legal professionals to incorporate culturally competent designs”).

<sup>194</sup> *Id.* at 370-71.

<sup>195</sup> See *supra* Part III.

like GPT-3 in the practice of law for them to be truly effective in governing that technology's use. The Comments to certain Rules should be updated to explicitly reflect the best practices that were discussed in Part III, much as Comment 8 to Rule 1.1 was updated in 2012 to specify duties of competence with respect to technology.<sup>196</sup> The Comments are the ideal place for amendments, as they were designed to help clarify and elucidate the existing rules, but not to impose any new obligations or restrictions. The preamble to the Model Rules notes that, while “the text of each Rule is authoritative,” “[t]he Comments are intended as guides to interpretation . . . .”<sup>197</sup>

Of course, as a threshold matter, it is important to acknowledge that the amendment process for the Model Rules has traditionally been lengthy and sometimes fraught. For example, the 2002 Model Rule amendments were first contemplated in 1997, when the ABA Ethics 2000 Commission was formed.<sup>198</sup> In the five years following the Commission's formation, that Commission “held fifty-one full days of meetings, held more than twelve public hearings, communicated regularly with its 250-member advisory council, consulted with special-interest groups, and made its discussion drafts and meeting minutes available on the internet.”<sup>199</sup> Only after the proposed rules were debated at two ABA meetings spaced several months apart were most of the amendments adopted.<sup>200</sup> Even once the Model Rules themselves are amended, it can take several years more for individual states to adopt them in whole or in part.

Thus, it could be a multi-year process for relevant comments to be added to the Model Rules, at which point we have moved on to GPT-4 (or GPT-5, etc.) and a new AI technology. Thus, the amended comments need to be specific enough to really guide lawyer behavior but not so specific that they are out of date as soon as they are published. The ABA has likely tried to thread this needle (and acknowledge the

---

<sup>196</sup> See *supra* Part III.A and accompanying footnotes.

<sup>197</sup> MODEL RULES OF PRO. CONDUCT, Preamble ¶ 21 (AM. BAR ASS'N 2020); see also *id.* ¶ 14 (“Comments do not add obligations to the Rules but provide guidance for practicing in compliance with the Rules.”); *id.* ¶ 15 (“The Comments are sometimes used to alert lawyers to their responsibilities under such other law.”).

<sup>198</sup> A.B.A., ETHICS 2000 COMMISSION EXECUTIVE SUMMARY, [https://www.americanbar.org/groups/professional\\_responsibility/policy/ethics\\_2000\\_commission/e2k\\_exec\\_summ/](https://www.americanbar.org/groups/professional_responsibility/policy/ethics_2000_commission/e2k_exec_summ/) [<https://perma.cc/F4KH-PUCH>] (noting that “[t]he Commission on Evaluation of the Rules of Professional Conduct was created in mid-1997”).

<sup>199</sup> Andrew F. Halaby & Brianna L. Long, *New Model Rule of Professional Conduct 8.4(g): Legislative History, Enforceability Questions, and a Call for Scholarship*, 41 J. LEGAL PROF. 201, 233 (2017).

<sup>200</sup> *Id.*

timeline for Rule amendment) through adopting resolutions rather than amending the rules. As noted previously, in 2019, the ABA passed a resolution “urging” lawyers and courts to “address the emerging ethical and legal issues related to the usage of artificial intelligence (“AI”) in the practice of law . . . .”<sup>201</sup> The ABA is correct to urge lawyers to pay attention to these important rules, but the resolution is a toothless tiger at this point. Neither the resolution nor the more detailed report that accompanied it “provide much in the way of specifics with regard to how courts and lawyers should address these emerging issues.”<sup>202</sup> Lawyers need more specificity and guidance to help ensure they are ethically deploying AI like GPT-3 in their legal practices.

#### A. *Amendments to Comment 8 to Rule 1.1*

Comment 8 to Rule 1.1 should be amended once more to be more specific about a lawyer’s duties with respect to the use of AI in the practice of law. In its current form, Comment 8 provides that lawyers should pay attention to a variety of changes to the legal profession, including “the benefits and risks associated with relevant technology,” and that they should engage in continuing legal education. Commentators have suggested that the language was kept intentionally broad, so that it would not have to be continually amended to deal with new technology.<sup>203</sup> But it is currently too vague to be useful. The amended comment should require that attorneys attend continuing legal education that is specifically addressed at the ethical use of AI in the practice of law.

Although it will be up to each state to determine how many CLE hours to require, mandating some amount of technology CLE will help signal to lawyers how important this topic is. Some states already require CLE credits in technology.<sup>204</sup> Lawyers in Florida, for example,

---

<sup>201</sup> AM. BAR ASS’N HOUSE OF DELEGATES, *supra* note 127, at 12.

<sup>202</sup> Bob Ambrogi, *ABA Votes to Urge Legal Profession to Address Emerging Legal and Ethical Issues of AI*, LAW SITES (Aug. 14, 2019), <https://www.lawsitesblog.com/2019/08/aba-votes-to-urge-legal-profession-to-address-emerging-legal-and-ethical-issues-of-ai.html> [<https://perma.cc/6GCW-WPFT>].

<sup>203</sup> See Baker, *supra* note 138, at 557, 560 (“The language of [the duty of competence] was left purposefully broad to account for technologies today, as well as technologies that have not yet been conceived. . . . The amended language found in Comment 8 is amorphous. This vague language was purposeful . . . .”).

<sup>204</sup> Medianik, *supra* note 128, at 1525; see also Victor Li, *Florida Supreme Court Approves Mandatory Tech CLE Classes for Lawyers*, A.B.A. J. (Sept. 30, 2016, 8:45 AM CDT), [https://www.abajournal.com/news/article/florida\\_supreme\\_court\\_approves\\_mandatory\\_tech\\_cles\\_for\\_lawyers](https://www.abajournal.com/news/article/florida_supreme_court_approves_mandatory_tech_cles_for_lawyers) [<https://perma.cc/F2JG-BS34>].



are required to take 3 hours of “approved technology program” CLE courses over a 3-year span.<sup>205</sup> An attorney who was involved in the effort to update Florida’s professional practice rules to include that requirement said it was relatively easy task to accomplish and “not as tough a sell as he and his subcommittee thought it would be.”<sup>206</sup>

Comment 8 should also address when and how an attorney can delegate some of their responsibility to be technically competent. Some state bar associations have already produced guidance on this topic. For example, the New York State Bar Association has promulgated social media ethics guidelines for attorneys.<sup>207</sup> Those guidelines provide guidance on attorneys’ technological competence, both with respect to social media use and beyond. With respect to delegation, the guidelines address electronic discovery and note that “[a]lthough a lawyer may not delegate his or her obligation to be competent, he or she may rely, as appropriate, on other lawyers or professionals in the field of electronic discovery and social media to assist in obtaining such competence.”<sup>208</sup>

#### B. Amendment to MCLE Model Rule

Alternatively, a focus on AI CLE could be accomplished through an amendment to the ABA’s Model Rule for Minimum Continuing Legal Education and Comments (“MCLE Model Rule”). Adopted in February of 2017 by the ABA’s House of Delegates, the MCLE Model Rule requires that lawyers take specialty CLE credits in three areas: (1) Ethics and Professionalism (average one credit per year); (2) Diversity and Inclusion (one credit every three years); and (3) Mental Health and Substance Use Disorders (one credit every three years). Adding a requirement for an annual credit in technology would be one way to help “give teeth” to the ABA’s Resolution 112, urging that lawyers educate themselves on AI.<sup>209</sup>

---

<sup>205</sup> RS. REGULATING FLA. BAR 6-10.3(b) (“Each member must complete a minimum of 33 credit hours of approved continuing legal education activity every 3 years . . . 3 of the 33 credit hours must be in approved technology programs.”).

<sup>206</sup> Li, *supra* note 204 (quoting attorney John M. Stewart as saying, “Throughout this entire process, we’ve gotten almost no pushback from lawyers . . . I think everyone recognized that lawyers could benefit from more education, both when it comes to technology and in general.”).

<sup>207</sup> N.Y. STATE BAR ASS’N, SOCIAL MEDIA ETHICS GUIDELINES (2019), <https://nysba.org/app/uploads/2020/02/NYSBA-Social-Media-Ethics-Guidelines-Final-6-20-19.pdf> [<https://perma.cc/VJN2-ZCUZ>].

<sup>208</sup> *Id.* at 5.

<sup>209</sup> See AM. BAR ASS’N OF DELEGATES, <https://www.americanbar.org/content/dam/aba/directories/policy/annual-2019/112-annual-2019.pdf> [<https://perma.cc/2LZ3-JSES>].

Admittedly, there has not been widespread adoption by the states of the MCLE Model Rule's requirement of specialty CLE credits.<sup>210</sup> Nonetheless, such an amendment would again signal the importance of the topic of AI-focused CLEs. There is a role for law schools to play here as well: "[L]aw schools can implement mandatory legal technology courses into their curricula or add the topic to the professional responsibility requirement."<sup>211</sup> Many schools now offer some version of "Artificial Intelligence and the Law." I developed and teach such a course at the West Virginia University College of Law. Students are introduced to the basics of artificial intelligence, including machine learning and algorithmic decision-making.<sup>212</sup> The course also covers the importance of explainability and interpretability in addressing bias in algorithms, and students spend a week discussing how AI will change the future of the legal profession.<sup>213</sup> Lawyers who are exposed to these issues as law students will be more receptive to any CLEs they take later on, and will be better positioned to ethically adopt (or reject) new technology as it evolves over the span of their own legal careers.

### C. Amendment to Comment 3 to Rule 5.3

In its current form, Comment 3 to Rule 5.3 provides that "a lawyer must make reasonable efforts to ensure that [services provided by a nonlawyer] are provided in a manner that is compatible with the lawyer's professional obligations." The Comment should be amended to address AI specifically, and should be even more explicit that lawyers have a duty to supervise AI systems themselves, and not just the technical support staff who may help select or run those systems. Toward that end, the Comment should provide more guidance on what reasonable efforts to supervise AI look like. Because "the duty of supervision goes hand-in-hand with the duty of competence for attorneys,"<sup>214</sup> a proposed amendment to Comment 3 of Rule 5.3 would be similar to the proposed amendment to Comment 8 to Rule 1.1. Put another way, in order to competently use an AI tool like GPT-3, you must supervise it. The Comment should make clear that such

---

<sup>210</sup> See ABA MCLE Chart, A.B.A. (Apr. 12, 2017), [https://www.americanbar.org/content/dam/aba/directories/policy/aba\\_model\\_rule\\_comparison\\_by\\_state\\_meet\\_model\\_rule\\_noted.pdf](https://www.americanbar.org/content/dam/aba/directories/policy/aba_model_rule_comparison_by_state_meet_model_rule_noted.pdf) [<https://perma.cc/DAU6-TZG5>] (comparing jurisdictional agreement with ABA model MCLE Rule).

<sup>211</sup> Medianik, *supra* note 128, at 1525.

<sup>212</sup> Syllabus, Professor Amy B. Cyphert, Artificial Intelligence and the Law (Summer 2021) (on file with author).

<sup>213</sup> *Id.*

<sup>214</sup> Schaefer, *supra* note 169, at 232.

---

supervision could involve associating with an expert, as the California Bar allows, provided that you supervise that expert.

The ABA should also consider issuing best practices that attorneys could follow when selecting and supervising experts in AI or other technology. This could be done through an expansion and refocusing of existing resources, such as those included in the ABA's Legal Technology Resource Center. That Center produces "publications, blog posts, webinars, and other free resources . . . to help[] lawyers identify opportunities, overcome obstacles, and understand how technology tools can improve their practices."<sup>215</sup> The Center could produce webinars devoted to selecting and supervising AI experts, allowing practitioners to gain insight and fulfill the CLE requirements proposed above. Providing materials through an existing resource center, and making sure that they are free or low cost, would be important. Otherwise, a requirement that lawyers participate in technology CLE credits may burden smaller firms, solo practitioners, and legal services organizations, as they are less likely to have in-house technology experts. In 2020, only 27% of solo practitioners reported to the ABA that they had some sort of technology training available to them, as opposed to 100% of attorneys from large (500+ lawyer) firms.<sup>216</sup> There is obviously a role for the ABA to play here.

#### D. A New Comment to Rule 8.4(g)

A new Comment should be added to Rule 8.4(g) that cautions lawyers from using AI systems without first understanding their propensity for bias. Lawyers are not going to be able to effectively remove bias from AI systems — a task that technical experts and data scientists have thus far been unable to accomplish.<sup>217</sup> But lawyers can and should be warned that failure to understand the potential for bias in these tools may lead to their misuse in the practice of law, that such misuse could lead to sanctions. The fact of bias in AI systems is well-documented in scholarly research, but not necessarily intuitive to most laypeople, who tend to

---

<sup>215</sup> Legal Technology Resource Center, A.B.A., [https://www.americanbar.org/groups/departments\\_offices/legal\\_technology\\_resources/](https://www.americanbar.org/groups/departments_offices/legal_technology_resources/) (last viewed July 13, 2021) [<https://perma.cc/H7DU-9PFH>].

<sup>216</sup> Mark Rosch, *ABA 2020 Tech Report*, A.B.A. (Nov. 16, 2020), [https://www.americanbar.org/groups/law\\_practice/publications/techreport/2020/techtraining/](https://www.americanbar.org/groups/law_practice/publications/techreport/2020/techtraining/) [<https://perma.cc/BDF6-D5WQ>].

<sup>217</sup> See Craig S. Smith, *Dealing with Bias in Artificial Intelligence*, N.Y. TIMES (Nov. 19, 2019), <https://www.nytimes.com/2019/11/19/technology/artificial-intelligence-bias.html> [<https://perma.cc/SL7V-NRUQ>] (discussing why bias in AI is such a difficult issue to address).

believe that technology is objective and neutral.<sup>218</sup> Cautioning lawyers about this bias, and making clear to them that it is professional misconduct to ignore it, will help focus attorneys on this topic.

Lawyers can learn more about the potential for bias in AI systems through a variety of ways. The mandatory CLEs described above could focus on AI and bias (the author of this Article led one for West Virginia attorneys on February 10, 2021). Lawyers could read any of a number of excellent law review articles that describe this.<sup>219</sup> But until lawyers realize the bias risks that AI tools like GPT-3 pose, they are unlikely to take these steps. A Comment to Rule 8.4(g) is a very important first step.

#### CONCLUSION

Like all AI, GPT-3 is neither inherently good nor inherently bad. Rather, it is full of both promise and peril. The technology may impact the practice of law for the better. For example, it could streamline the drafting process, reducing fees for clients. Or it could impact the practice of law for the worse, spewing toxic language and perpetuating existing biases. The impacts are far from predetermined and are difficult to predict. It is easy to see, however, that the more ethical oversight lawyers exercise, the more they think critically about the technology and how or if to use it, the more they understand the inherent limitations and downsides, then the higher the likelihood that the technology will represent a net positive for lawyers and their clients. Lawyers need clear direction on how to ethically use GPT-3, and they need an incentive to follow that direction and the support and resources to do so. GPT-3 may soon be replaced by GPT-4, or by whatever the next “wonder tech” is. But AI as a whole will not be replaced in the practice of law, and its impact will only grow.

---

<sup>218</sup> See, e.g., Cyphert, *supra* note 63, at 473 (“It is tempting to think of any artificial intelligence, including an algorithm, as neutral and objective. Laypeople without technical expertise can be especially vulnerable to placing too much faith in algorithmic outcomes.”).

<sup>219</sup> See, e.g., Solon Barocas & Andrew D. Selbst, *Big Data’s Disparate Impact*, 104 CALIF. L. REV. 671 (2016) (discussing bias and big data); David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 UC DAVIS L. REV. 653 (2017) (discussing bias and machine learning); Mayson, *supra* note 59 (discussing bias in criminal justice strategies).