LAWYERING IN THE AGE OF ARTIFICIAL INTELLIGENCE

Jonathan H. Choi,[*] Amy B. Monahan[**] & Daniel Schwarcz[***]

We conduct the first randomized controlled trial of AI assistance's effect on human legal analysis. We randomly assigned sixty students at the University of Minnesota Law School each to complete four separate legal tasks (drafting a complaint, a contract, a section of an employee handbook, and a client memo), either with or without the assistance of GPT-4, after receiving training on how to use GPT-4 effectively. We then blind-graded the results and tracked how long the students took on each task.

We found that access to GPT-4 slightly and inconsistently improved the quality of participants' legal analysis but induced large and consistent increases in speed. The benefits of AI assistance were not evenly distributed: in the tasks on which AI was the most useful, it was significantly more useful to lower-skilled participants. On the other hand, AI assistance reduced the amount of time that participants took to complete the tasks roughly uniformly regardless of their baseline speed. In follow up surveys, we found that participants reported increased satisfaction from using AI to complete legal tasks and that they correctly predicted the tasks for which GPT-4 would be most helpful.

These results—which will likely serve as a lower-bound estimate on AI's capacity to improve the efficiency of legal services—have important normative implications across the future of lawyering. For law schools, they suggest the importance of deliberately and holistically assessing when and how law students are trained to use AI. For lawyers and judges, they suggest that the time to embrace AI is now, though the contours of what that will mean can and should vary significantly by practice area, task, and the stakes of the underlying matters. And for purchasers of legal services, our results suggest that it is time to reconsider what types of legal matters should be sent to outside counsel rather than handled in-house, and how matters that are handled externally are managed and billed.

---

[*] Professor of Law, University of Southern California Gould School of Law.

[**] Distinguished McKnight University Professor, University of Minnesota Law School.

[***] Fredrikson & Byron Professor of Law, University of Minnesota Law School.

Contents

INTRODUCTION

Rapid improvements in the performance of artificial intelligence (AI) models over the past two years have triggered excitement and trepidation about the future of lawyering.[1] Will AI replace human lawyers, or help them become more efficient? Is AI more useful at some legal tasks than others, or more useful to some lawyers than others? And how should lawyers, law students, and law schools adjust to the increasing availability and development of AI?

Studies to date offer only limited answers to these questions, as they have focused on AI's ability to conduct legal analysis on its own, rather than its ability to assist humans, which we consider the more plausible use case for the foreseeable future. In addition, for reasons of convenience, existing studies have generally applied GPT-4 to exams (like law school exams and the bar exam)[2] and have suffered from methodological limitations, like non-blind grading of results[3] or imperfectly matched treatment and control groups.[4]

To better understand how AI will affect the lawyers of the future, and what should be done about that possibility now, we conduct the first randomized controlled trial of the effect of AI assistance on human legal analysis. We randomly assigned sixty students at the University of Minnesota Law School to complete four separate legal tasks (drafting a complaint, a contract, a section of an employee handbook, and a client memo) either with or without the assistance of GPT-4, after receiving training on how to use GPT-4 effectively. We then blind-graded the results and tracked how long the students took on each task.

---

[1] Roger E. Barton, *How Will Leveraging AI Change the Future of Legal Services?,* Reuters, Aug. 23, 2023; Daniel Farrar, *To Future-Proof Their Firms, Attorneys Must Embrace AI*, Forbes, July 13, 2023; Steve Lohr, *A.I. is Coming for Lawyers, Again*, N.Y. Times, April 10, 2023; John Villasenor, *How AI Will Revolutionize the Practice of Law,* Brookings Institute Commentary, March 20, 2023, https://www.brookings.edu/articles/how-ai-will-revolutionize-the-practice-of-law/.

[2] *See infra* __.

[3] *See* [[GPT-4 passes the bar exam, others]].

[4] *See* [[Choi & Schwarcz 2023]].

We found that access to GPT-4 only slightly improved the quality of participants' legal analysis, with improvements that were small in magnitude and inconsistent across tasks (+0.17, +0.24, +0.07, -0.07 on a 4.0 grading scale). However, we found that AI assistance induced large and consistent declines in the amount of time taken to complete tasks (-24.1%, -32.1%, -21.1%, -11.8%). The benefits of AI assistance were not evenly distributed; in the tasks on which AI was the most useful, it was significantly more useful to lower-skilled participants (judged by their scores on tasks for which they did not have AI assistance). On the other hand, AI assistance reduced the amount of time that participants took to complete the tasks roughly uniformly regardless of their baseline speed.

We also surveyed participants on their perceptions of how access to GPT-4 impacted their work on legal tasks. We found that (again for the tasks on which GPT-4 was most useful) participants reported increased satisfaction from using it. Participants also correctly understood GPT-4's strengths and weaknesses, reporting that they expected the improvements in speed were greater than the improvements in quality and correctly identifying the tasks at which GPT-4 was most useful. This suggests that although the benefits from AI use may be uneven, participants generally correctly perceived the tasks at which it was most useful.

These results——which will likely serve as a lower-bound estimate on AI's capacity to improve the efficiency of legal services——have important normative implications for actors across the legal services industry. For law schools, they suggest the importance of deliberately and holistically assessing when and how law students are trained to use AI, and when and how access to that tool should be limited. For lawyers and judges, our results suggest that the time to embrace AI is now, though the contours of what that will mean can and should vary significantly by practice area, task, and the stakes of the underlying matters. Purchasers of legal services also should pay close attention to our results, reconsidering what types of legal matters should be sent to outside counsel rather than handled in-house, and how matters that are handled externally are managed and billed.

I. BACKGROUND

This Part briefly reviews both the evolution of legal technology and the state of the scholarly literature on AI and lawyering and other knowledge-based tasks.

Fifty years ago, at the beginning of what we might think of as the modern era of legal technology, the first legal databases were

introduced.[5] Over the next decades, innovations such as email, document management systems, billing software, e-discovery systems, and online dispute resolution platforms were widely adopted and helped shape practice patterns. In addition, tech-based "disrupters" such as Rocket Lawyer, Legal Zoom, and Trust & Will entered the market, offering an online, often automated, solution for the drafting of common legal documents.[6]

Historically, these major legal tech innovations have improved lawyer efficiency rather than fundamentally altering the core skills needed to be an effective lawyer.[7] For example, a lawyer with access to an easily searchable legal database can complete legal research in much less time than would be possible if they needed to search through hard copy indices. But the skill involved in analyzing and applying cases and statutes remains fundamentally the same.[8] Similarly, e-discovery tools allow lawyers to automate the search function in discovery[9] but cannot provide the knowledge necessary to identify what must be produced and what is protected by privilege.

Even before the recent wave of progress in generative AI tools like ChatGPT, the rise of AI in legal tech was disrupting this historical pattern. To illustrate, recent AI tools like predictive coding in e-discovery systems have become increasingly prominent in recent years. These tools allow a lawyer to code a sample of discovery documents, which are then used to by an algorithm to identify other relevant documents.[10] To a certain degree, tools such as these actually displace an attorney's work.[11]

---

[5] William G. Harrington, *A Brief History of Computer-Assisted Legal Research*, 77 LAW LIBR. J. 543, 553 (1985).

[6] *See* Susan Saab Fortney, *Online Legal Document Providers and the Public Interest: Using a Certification Approach to Balance Access to Justice and Public Protection*, 72 Okla. L. Rev. 91, 93 (2019).

[7] Mark Fenwick et. al., Legal Education in the Blockchain Revolution, 20 Vand. J. Ent. & Tech. L. 351, 357 (2017).

[8] *See* Raymond H. Brescia et al., *Embracing Disruption: How Technological Change in the Delivery of Legal Services Can Improve Access to Justice*, 78 Alb. L. Rev. 553, 568 (2014).

[9] *See* John O. McGinnis & Russell G. Pearce*, The Great Disruption: How Machine Intelligence Will Transform the Role of Lawyers in the Delivery of Legal Services,* 82 Fordham L. Rev. 3047-48 (2014).

[10] *See* id.

[11] *See* Maura R. Grossman & Gordon V. Cormack, *Quantifying Success: Using Data Science to Measure the Accuracy of Technology-Assisted Review in Electronic Discovery, in* DATA DRIVEN LAW: DATA ANALYTICS AND THE NEW LEGAL SERVICES 127, 150-51 (Ed Walters ed.,

With each new innovation, lawyers have typically fretted about the implications for the legal profession and lawyer jobs.[12] If technology allowed the same work to be done in less time,[13] or could replace lawyers altogether for certain tasks,[14] it was feared that there would be fewer jobs available for lawyers. Of course, others championed at least some of these advances as having the potential to lower legal fees and therefore increase access to legal services.[15]

2019) (finding that these "technology-assisted review" systems in e-dicovery provided "significantly superior precision" compared to manual review). *But see* Emily S. Taylor Poppe, *The Future Is Bright Complicated: Ai, Apps & Access to Justice*, 72 Okla. L. Rev. 185, 189 (2019) (arguing that displacement concerns are less significant when it comes to tasks that were already subject to outsourcing).

[12] See, e.g., RICHARD SUSSKIND & DANIEL SUSSKIND, THE FUTURE OF THE PROFESSIONS: HOW TECHNOLOGY WILL TRANSFORM THE WORK OF HUMAN EXPERTS (2015); Daniel Martin Katz, *Quantitative Legal Prediction--Or--How I Learned to Stop Worrying and Start Preparing for the Data-Driven Future of the Legal Services Industry*, 62 Emory L.J. 909 (2013); McGinnis & Pearce, *supra* note x; Dana A. Remus & Frank Levy, *Can Robots Be Lawyers: Computers, Lawyers, and the Practice of Law*, 30 GEO J. LEGAL ETHICS 501 (2017); Tanina Rostain, *Robots versus Lawyers: A User-Centered Approach*, 30 GEO J. LEGAL ETHICS 559 (2017); Sean Semmler & Zeeve Rose, *Artificial Intelligence: Application Today and Implications Tomorrow*, 16 Duke L. & Tech. Rev. 85 (2017); Harry Surden, *Machine Learning and Law*, 89 Wash L. Rev. 87 (2014); David C. Vladeck, *Machines without Principals: Liability Rules and Artificial Intelligence*, 89 Wash. L. Rev. 117 (2014); John Markoff, *Armies of Expensive Lawyers, Replaced by Cheaper Software*, N.Y. Times (Mar. 4, 2011), https://www.nytimes.com/2011/03/05/science/05legal.html; James E. Moliterno, The American Legal Profession in Crisis: Resistance and Responses to Change 208 (2013).

[13] *See* Mark Fenwick et. al., *Legal Education in the Blockchain Revolution*, 20 Vand. J. Ent. & Tech. L. 351, 357 (2017)

[14] Christopher A. Suarez, *Disruptive Legal Technology, Covid-19, and Resilience in the Profession*, 72 S.C. L. Rev. 393, 404 (2020).

[15] *See, e.g.*, Susskind & Susskind, *supra* note x, at 66-67; McGinnis & Pearce, supra note x; Raymond H. Brescia et. al., *Embracing Disruption: How Technological Change in the Delivery of Legal Services Can Improve Access to Justice*, 78 Alb. L. Rev. 553, 553 (2015); Elinor R. Jordan, *Point, Click, Green Card: Can Technology Close the Gap in Immigrant Access to Justice?*, 31 Geo. Immigr. L.J. 287 (2017); Elliott Vinson & Samantha A. Moppett, *Digital Pro Bono: Leveraging Technology to Provide Access to Justice*, 92 St. Johns L. Rev. 551 (2018).

Similar dynamics exist in recent discussions of how increasingly capable large languages models (LLMs) like GPT-4[16] will impact the legal profession. At the same time, LLMs like GPT-4 seems to represent a qualitatively different type of technological advance from those that came before. As a result, many have speculated that these LLMs will lead to true revolution in the practice of law,[17] radically changing market demand for human lawyers.[18]

Yet, despite these sizeable questions and concerns, relatively little is known empirically about AI's capacity to displace lawyers or even capably assist lawyers at lawyering tasks. To date, the best information we have is found in studies of GPT-4's performance on law school examinations, bar examinations, and in answering discrete legal questions. Other non-empirical research considers the ethical implications of using such technology in the practice of law,[19] how artificial intelligence may change the skills needed to be a successful lawyer,[20] and the manner in which law firms may begin to compete on the basis of technological expertise.[21]

Studies examining GPT's proficiency on legal exams have found that its performance varies widely depending on the type of exam and

---

[16] *See, e.g.,* Erin Mulvaney & Laura Webber, *End of the Billable Hour? Law Firms Get On Board With Artificial Intelligence*, Wall St. J. May 11, 2023.

[17] Even before the advent of large language model AI, some "legal futurists" were envisioning such transformation. *See, e.g.,* Benjamin Alarie, *The Path of the Law: Towards Legal Singularity*, 66 U. Toronto L.J. 443, 445 (2016) (describing the "legal singularity" that will occur when "the accumulation of a massive amount of data and dramatically improved methods of inference make legal uncertainty obsolete").

[18] The impact of generative AI on the labor market is certainly not limited to the legal profession. *See, e.g.,* Tyna Eloundou et al., *GPTs and GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models*, arXiv:2303.10130v5 (Aug. 22, 2023), https://arxiv.org/pdf/2303.10130.pdf.

[19] *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 (2018).

[20] *See, e.g.,* Alyson Carrel, *Legal Intelligence Through Artificial Intelligence Requires Emotional Intelligence: A New Competency Model for the 21st Century Legal Professional,* 35 Ga. St. U. L. Rev. 1153 (2019); Christopher A. Suarez, *Disruptive Legal Technology, Covid-19, and Resilience in the Profession*, 72 S.C. L. Rev. 393 (2020).

[21] Bruce A. Green & Carole Silver, *Technocapital@biglaw.com*, 18 Nw. J. Tech. & Intell. Prop. 265 (2021).

prompting methodology used. One study found that GPT-4 alone performed in the 90th percentile on the Uniform Bar Examination[22] (although scholars have subsequently raised methodological doubts about this claim[23]). In another study evaluating AI-generated answers to law school exam questions, researchers found that although exams drafted by GPT-3.5 often included solid explanations of basic legal rules and strong organization and composition, they also often struggled to identify relevant issues and tended to only superficially apply rules to facts as compared to real law students.[24] Perhaps most interestingly, a

---

[22] Daniel Martin Katz et al., GPT-4 Passes the Bar Exam (Apr. 5, 2023) (unpublished manuscript) (on file with authors). This result extended both to the multiple-choice portion of the exam as well as to the open-ended essay components of the exam. *Id.* at *2. Although the authors did not use any prompt-engineering strategies to generate multiple choice answers, they slightly modified essay questions by presenting each sub-question in an independent prompt, and by "lightly correcting the language" in the prompt so that it formed a complete sentence. *Id.* at *7.

[23] Eric Martínez, Re-Evaluating GPT-4's Bar Exam Performance (June 12, 2023) (unpublished manuscript) (on file with authors) (discussing potential methodological issues with the initial finding that GPT-4 surpassed the bar exam score of 90% of human test takers). In addition, the authors in the Katz et al. study did not grade GPT-4's performance blind and did not have experience grading bar exams, raising concerns about subjective bias in evaluation.

[24] Jonathan H. Choi, Kristin E. Hickman, Amy B. Monahan, & Daniel Schwarcz, *ChatGPT Goes to Law School*, 71 J. Leg. Ed. 387 (2022) (testing the performance of GPT-3.5 alone on law school exams). *See also* Andrew Blair-Stanek et al., GPT-4's Law School Grades: Con Law C, Crim C-, Law & Econ C, Partnership Tax B, Property B-, Tax B (May 24, 2023) (unpublished manuscript) (on file with authors); Margaret Ryznar, *Exams in the Time of ChatGPT*, 80 Wash. & Lee L. Rev. Online 305 (2023) (reporting mixed results).

In other disciplines, GPT has been found to be a proficient and sometimes superior test taker as compared to humans. *See* Harsha Nori et al., Capabilities of GPT-4 on Medical Challenge Problems (Apr. 12, 2023) (unpublished manuscript) (on file with authors) (finding that GPT-4, without any specialized prompting passes a range of medical exams and out-performs both ChatGPT and LLM models specifically fine-tuned on medical knowledge); John C. Lin et al., *Comparison of GPT-3.5, GPT-4, and Human User Performance on a Practice Ophthalmology Written Examination*, Eye, May 8, 2023 ("GPT-4 but not GPT-3.5 achieved the passing threshold for a practice ophthalmology written examination");

later study examining GPT-4 assistance on law school exams (i.e., where some study participants used GPT-4 to help generate exam answers, but then reviewed those answers and edited them as they felt appropriate) found that such assistance boosted the scores of lower-performing students but had no effect or a slightly negative effect on the performance of top students.[25]

Outside of the exam context, little evidence exists on how access to LLM tools like GPT-4 might impact lawyers' or law students' abilities to complete legal tasks. Tax scholars have tested GPT-4's ability to answer questions about federal tax law, generally finding low accuracy with basic prompting (roughly 30% in two separate studies) to 70%-90% accuracy with significant human assistance (particularly prompting with hand-selected correct sources).[26] Many scholars have anecdotally tested GPT's capabilities, including a series of YouTube videos that illustrate GPT-4's capabilities in various legal contexts.[27] Thee anecdotal reports find, for example, that with good prompting, GPT-4 is able to accurately apply copyright law, although its performance falters on more difficult legal analysis.[28]

In areas other than law, we see the same general focus on exam performance rather than studies of realistic tasks. And as with law, the exam results are mixed. Whereas exams generated by ChatGPT were

---

Rohaid Ali et al., Performance of ChatGPT and GPT-4 on Neurosurgery Written Board Examinations, Neurosurgery p. 10 (2023) (finding that both GPT-4 and GPT-3.5 pass neurosurgery practice board exams at rates comparable to neurosurgery residents); Hanmeng Liu et al., Evaluating the Logical Reasoning Ability of ChatGPT and GPT-4 (May 5, 2023) (unpublished manuscript) (on file with authors); Vinay Pursnani, Yusuf Sermet & Ibrahim Demir, Performance of ChatGPT on the US Fundamentals of Engineering Exam: Comprehensive Assessment of Proficiency and Potential Implications for Professional Environmental Engineering Practice (Apr. 20, 2023) (unpublished manuscript) (on file with authors).

[25] Choi & Schwarcz

[26] John Net et al., *L*arge Language Models as Tax Attorneys: A Case Study in Legal Capabilities Emergence, 381 PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A: MATHEMATICAL, PHYSICAL AND ENGINEERING SCIENCES (forthcoming 2023); Andrew Blair-Stanek, Nils Holzenberger, & Benjamin Van Durme, OpenAI Cribbed Our Tax Example, But Can GPT-4 Really Do Tax?, 180 Tax Notes Fed. 1101, 1105 (2023).

[27] https://www.youtube.com/@harrysurden3116

[28] https://www.youtube.com/watch?v=nqZcrhR8yPU

rated as "outstanding" in economics[29], they achieved more middling results in computer programming and medical education,[30] and "unsatisfactory" results in fields like mathematics and psychology.[31] Common problems with ChatGPT-drafted exams included inaccurate, unreliable, and outdated information.[32] These studies vary significantly in the methods they use to test LLM performance. Some test the performance of AI acting alone, where a question or prompt is entered into an LLM and its answer is evaluated without modification. Other studies examine the value of AI *assistance*, where a human subject uses an LLM on various tasks or subtasks and then reviews, edits, or otherwise refines those results to produce a final work product.

Outside of the exam setting, a small number of studies have evaluated how AI can improve human performance at non-legal professional writing tasks.[33] One study found that giving college-educated professionals access to GPT-3.5 substantially improved their performance at a variety of writing tasks, with the greatest gains going

---

[29] Wayne Geerling et al., *ChatGPT has Aced the Test of Understanding in College Economics: Now What?*, 68 Amer. Econ. 233 (2023) (finding that GPT ranked in the 91st percentile for Microeconomics and the 99th percentile for Macroeconomics when compared to college students taking the Test of Understanding in College Economics).

[30] Tiffany H. Kung et al., *Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models*, PLOS Digital Health, Feb. 9, 2023 (reporting that ChatGPT performed "at or near the passing threshold" on the United States Medical Licensing Exam). *See also* Peter Lee, Sebastien Bubeck, & Joseph Petro, *Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine*, 388 New Eng. J. Med. 1233 (2023).

[31] Se Chung Kwan Lo, *What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature* ,13 Educ. Sci. 410 (2023); see also ChatGPT Could Be a Stanford Medical Student, a Lawyer, or a Financial Analyst. Here's a List of Advanced Exams the AI Bot Has Passed So Far. Available online: https://www.businessinsider.com/list-here-are-the-exams-chatgpt-has-passed-so-far-2023-1 March 10, 2023.

[32] See Lo.

[33] There are some recent papers that evaluate how access to generative AI can improve professionals' ability to perform non-writing tasks, like computer coding. *See* Sida Peng et al., The Impact of AI on Developer Productivity: Evidence from GitHub Copilot (Feb. 13, 2023) (unpublished manuscript) (on file with authors). None of these studies evaluate how more sophisticated prompting techniques can impact results.

to the least-skilled workers.[34] On the other hand, other empirical work has suggested that human use of AI to assist with certain tasks can undermine humans' incentives to take care.[35]

One of the most extensive studies of AI-assistance in knowledge-intensive work examined the effect of AI-assistance on a range of work tasks common within the field of high-level management consulting. [36] The results show that AI is remarkably capable of increasing both quality and productivity on certain types of tasks but not others, even where the tasks are considered of similar difficulty. Specifically, consultants completing a series of tasks that involved conceptualizing and developing new product ideas significantly improved both the quality and speed of their work with the assistance of AI.[37] Where consultants were working

---

[34] Shakked Noy & Whitney Zhang, *Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence*, 381 SCIENCE 187, 187 (2023). To reach this conclusion, the experimenters recruited over 400 participants in five professional categories: grant writers, consultants, data analysts, human resource professionals, and managers. Participants were then tasked with completing two short writing assignments comparable to those they would complete in their professional settings, such as drafting press releases, short reports or emails. After completing the first writing assignment, half of the participants were given access to ChatGPT for the second writing assignment. The study found that participants who were provided with access to ChatGPT completed their writing tasks faster and produced higher quality work than participants who were not provided access to this tool. Moreover, the participants who performed relatively poorly on the initial task (which took place prior to being instructed how to use ChatGPT) disproportionately benefited from access to AI, receiving both higher quality scores and taking decreased amounts of time to complete their writing task. By contrast, access to ChatGPT did not improve the quality of work for participants who scored well in the initial writing task, though it did increase the speed at which they could produce that work.

[35] Fabrizio Dell'Acqua, Falling Asleep at the Wheel: Human/AI Collaboration in a Field Experiment on HR Recruiters, available at

[36] Fabrizio Dell'Acqua et al., Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality, Harvard Business School Working Paper 24-103, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4573321

[37] Id. at 9-10. *See also* Karan Girotra et al., *Ideas are Dimes a Dozen: Large Language Models for Idea Generation in Innovation*, Working Paper, July 10, 2023),

on problem-solving tasks that required the synthesis of quantitative data and qualitative information from interviews, AI provided much less of a boost.[38] Further, the greatest gains on both tasks were seen in the group that not only used AI assistance, but were also trained in effective prompt engineering.[39] The study also found, consistent with studies conducted by Choi & Schwarcz and Noy & Zhang, that the most significant beneficiaries of AI assistance were lower-skilled participants.[40] However, in contrast to Choi & Schwarcz, the study found performance improvements even among those in the top half of skill rankings.[41] While quality and productivity improved in all groups utilizing AI, the study found that on tasks involving creativity, those using the assistance of AI showed less variability in ideas than among those working without AI.[42] Researchers also found that participants who blindly adopted AI outputs suffered a decrease in performance compared to those not using AI assistance at all.[43]

In sum, the literature to date suggests that AI holds real promise to effectively assist with lawyering and other knowledge-based tasks, but also comes with some well-documented shortcomings. GPT-4 and other LLMs sometimes hallucinate sources and sometimes fail to interpret sources accurately. In addition, there are indications from several studies that the lowest-skilled workers benefit the most from AI assistance, with AI providing no benefit to or even possibly a negative effect on the performance of highly skilled humans.

Our study aims to move the literature forward by evaluating the effect of GPT-4 assistance, in terms of both quality and efficiency, on four

---

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4526071 (finding that GPT-4 can generate ideas faster and cheaper than college students at an elite university).

[38] Id. at 13-15.

[39] Id. at 10, 15.

[40] Id. at 11 (finding a 43% increase in performance among those ranked in the bottom half of skill level).

[41] Id. at 11 (finding a 17% increase in performance among to—half-skill subjects).

[42] Id. at 12. *See also* Leonard Boussioux et al., *The Crowdless Future? How Generative AI is Shaping the Future of Human Crowdsourcing*, Harvard Business School Working Paper 2402995 (2023), https://www.hbs.edu/ris/Publication%20Files/24-005_6a4d999b-82b8-496d-af15-6104c8876e74.pdf (similarly finding that GPT-4 may decrease some forms of creativity and novelty compared to purely human outputs).

[43] Id. at 17.

different lawyering tasks that are representative of the types of tasks a junior attorney might be asked to perform.

II. METHODOLOGY

We recruited students from the University of Minnesota Law School in April 2023 to participate in our study over Summer of 2023.[44] Well over 100 students expressed interest in participating in the study.[45] We initially enrolled the first sixty such volunteers and placed the remaining volunteers on a waitlist.[46] Over the duration of the study, 22 of the participants dropped out because they were unable to complete the entirety of the experiment; as they did so, we replaced them with new participants from the waitlist to ensure that we achieved roughly the target number of sixty study participants. Ultimately 59 students completed the experiment.

During the enrollment process, we gathered basic information about study participants, including their first-year law school GPA and their anticipated graduation year.[47] We then randomly sorted these participants into two thirty-person groups and confirmed that these two groups were roughly balanced with respect to graduation year and first-semester law school grade point average, as described in Section A of the Appendix.

---

[44] The University of Minnesota Law School is one of the top law schools in the country, currently ranked 16th in the U.S. News ranking of law schools. *2023 Best Law Schools*, U.S. NEWS, https://www.usnews.com/best-graduate-schools/top-law-schools/law-rankings (last visited Aug. 5, 2023).

[45] One of the co-authors sent a recruiting email to the entire University of Minnesota Law School student body in April, 2023. The email explained that we were recruiting "current JD students, including class of 2023 graduates, for participation in a study that examines the use of artificial intelligence tools, specifically GPT- 4, to assist with basic lawyering tasks." To participate in the study, students or graduates would need to be available to work for up to 15 hours total during June 2023. The email also noted that the work could be completed remotely and on participants' own time-schedules and that participants who completed the study receive $300 in compensation for their time. .

[46] This experimental design was approved by the University of Minnesota's IRB. Participants agreed to participate after reviewing and agreeing to an IRB-approved consent form.

[47] We also collected [information]

Study participants completed the experiment remotely, on their own schedule, from June to early August of 2023. Initially, they completed three online training modules that we developed and taught on how to use GPT-4 effectively in legal analysis.[48] Doing so required students to watch approximately two hours of training videos and to complete several short exercises using GPT-4. The training included both general techniques on how to prompt GPT-4 effectively (for example, by breaking down legal analysis into pieces and supplying relevant legal rules or sources) and how to use it specifically in litigation and transactional settings. It focused on how to apply active lawyering skills while using AI, rather than mechanically relying on the output of GPT-4. For example, we instructed participants to first assess assignments on their own before using GPT-4 to generate answers.[49] Additionally, the training required participants to practice these skills by using GPT-4 to answer sample problems. Section C of the Appendix provides additional information about the training materials used.

After completing the training, the participants then completed four basic lawyering tasks, representing a range of common tasks for entry-level lawyers.

The first assignment involved drafting a complaint for a fictional client to be filed in federal court on the basis of Section 1983, intentional interference with a business relationship, and malicious prosecution. Participants were not required to perform independent legal research for this task; they were provided with the elements of each cause of action in order to draft the complaint. The maximum time permitted for this task was five hours.

The second task required drafting a simple contract between a homeowner and housepainter. Participants were provided with the key terms of the contract and instructed to write the contract in plain English with a length not to exceed two pages. Participants were instructed to spend no more than two hours on this task.

The third assignment required participants to draft a short section of an employee handbook that explains employees' rights under federal and state (Minnesota) law to take breaks in order to pump breastmilk for a child. This task required legal research, as participants were not provided with the relevant statutes. Participants were

[48] This training drew heavily on previous work by two of us. See AI Tools for Lawyers: A Practical Guide. Daniel Schwarcz & Jonathan H. Choi, *AI Tools for Lawyers: A Practical Guide*, 108 MINN. L. REV. HEADNOTES (forthcoming 2023).

[49]

instructed to limit their work product to a single page and spend no more than one hour on this task.

The fourth and final task involved a fictional client with a potential product liability issue—namely, whether the client should be advised to place a warning label on a product when the product contains an allergen. The task required participants to read four provided cases but did not require independent legal research to complete. Each participant drafted a legal memorandum to the client offering legal analysis and advice on how best to proceed. Participants were instructed to spend no more than five hours on this task.

In addition to submitting their work product, each participant was asked to track the time they spent completing each task, and that time allocation was recorded separately from the work product so that it would not influence grading in any way.

Participants were compensated at a flat rate for their study participation in order to prevent participants from spending more time than necessary on a task in order to maximize their compensation. Participants also received the following instructions for each task:

> You should approach the assignment as if you are a junior attorney who has been asked to produce work for a fee-sensitive client. While you can take up to the maximum time allotment to complete the task, you should stop working at the point where you would feel comfortable submitting your work product to a supervising attorney, given that your client would prefer to minimize the amount they pay for your work product. If you reach the end of the maximum time allocation and have not finished, you should simply turn in the work product you were able to produce within the allotted time. Do not spend any more than the maximum time on any assignment.

The participants were divided between two groups, Group A and Group B. Each participant, whether assigned to Group A or Group B, was required to complete all four tasks. However, each group was instructed to use the assistance of GPT-4 on two of the four tasks, and to refrain from using GPT-4 or any other type of AI for the remaining two tasks. Specifically, Group A used GPT-4 for the contract drafting and complaint drafting tasks, while Group B used GPT-4 for the employee handbook and client memo tasks.

To provide access to GPT-4 to participants, we created a central ChatGPT "clone" website using the GPT-4 API, and gave students access

to that website. This clone website had a nearly identical user interface and used the same system prompt as the real ChatGPT Plus with GPT-4.

After all study participants had completed the four tasks in the experiment, we graded all participant work product anonymously, with no knowledge of participant identity or GPA, GPT use, or time spent on task. Grades were assigned based on grading standards and norms at the University of Minnesota Law School, where each study investigator has taught, but were not adjusted or "curved" in any manner. Each task was graded in its entirety by a single investigator using a pre-determined grading rubric to help ensure consistency.

At the completion of the experiment, all participants were asked to take an anonymous survey regarding their experience. Although the survey was anonymous on a per-respondent basis, we tracked responses separately for Groups A and B, allowing us to register how each group felt on average about their respective assignments. We pre-registered our methods and hypotheses prior to analyzing our results; the pre-registration statement is archived with the Open Science Foundation.[50]

III. RESULTS

Overall, we found that access to AI caused little average improvement on the quality of output in lawyering tasks but a substantial increase in speed of completion. However, the boost in quality from AI assistance depended on baseline: participants who had the worst performance without assistance from GPT-4 received the largest benefits, with little benefit to participants who were capable of producing high-quality work on their own. In contrast, the boost to speed was largely consistent among participants. When surveyed on their impressions, participants reported positive impressions of the AI, including positive reviews for the AI's impact on both speed and quality. Respondents indicated that their ability to use AI improved over the course of the experiment and that they were more likely to use AI tools in the future as a result of the experiment. Finally, respondents accurately assessed the tasks for which AI was most helpful even before having their grades revealed to them.

Table __ below shows statistics for the grades received and time taken for each task.[51] It shows that the differences are relatively small in magnitude. Access to GPT-4 had the largest positive effect for contract

[50] https://osf.io/5yzj3/

[51] All confidence intervals in this Article were generated using empirical bootstraps with 10,000 iterations.

drafting, where the difference in grade it generated was approximately two thirds of the difference between a B and a B+. The results also show substantial variation between tasks. On the client memo and EE handbook task, respondents saw, on average, a near zero effect on performance from using GPT-4.

Table 1: Effect of GPT-4 on Performance (Grade on 4.0 Scale)

	No GPT (Std. Dev.)	With GPT (Std. Dev.)	Difference (95% CI)
Complaint Drafting	3.14	3.31	0.17
	(0.59)	(0.50)	(-0.03, 0.37)
Contract Drafting	3.00	3.24	0.24
	(0.56)	(0.40)	(0.07, 0.41)
EE Handbook	3.20	3.26	0.07
	(0.41)	(0.39)	(-0.07, 0.21)
Client Memo	2.92	2.85	-0.07
	(0.69)	(0.76)	(-0.34, 0.18)

Figures __ through __ below depict the simple distribution of grades on tasks for groups with and without AI assistance. These Figures are density plots, presenting the number of participants (on the *y*-axis) who received each grade (on the *x*-axis).[52] Figures __ through __ in the Appendix show the bootstraps for the difference in means for groups with and without access to GPT, showing that only contract drafting showed a statistically significant increase in performance at the 95% level.

[52] All figures in this Article were generated using the SciPy package in Python. Density plots were generated using Gaussian Kernel Density Estimation using the gaussian_kde package in SciPy, applying Scott's rule of thumb to determine bandwidth. __.

Figure 1: Quality Distributions with and Without AI—Complaint Drafting

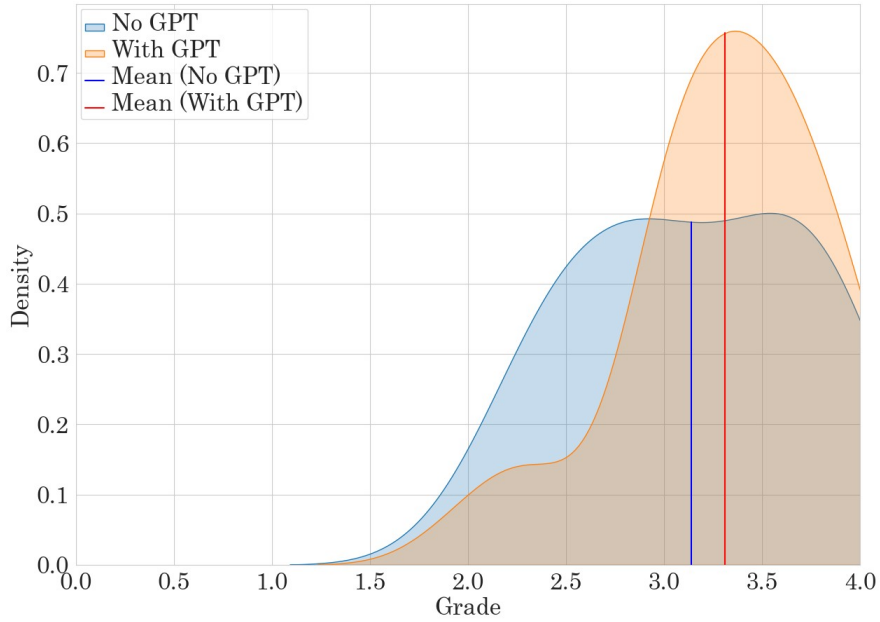


Figure 2: Quality Distributions with and Without AI—Contract Drafting

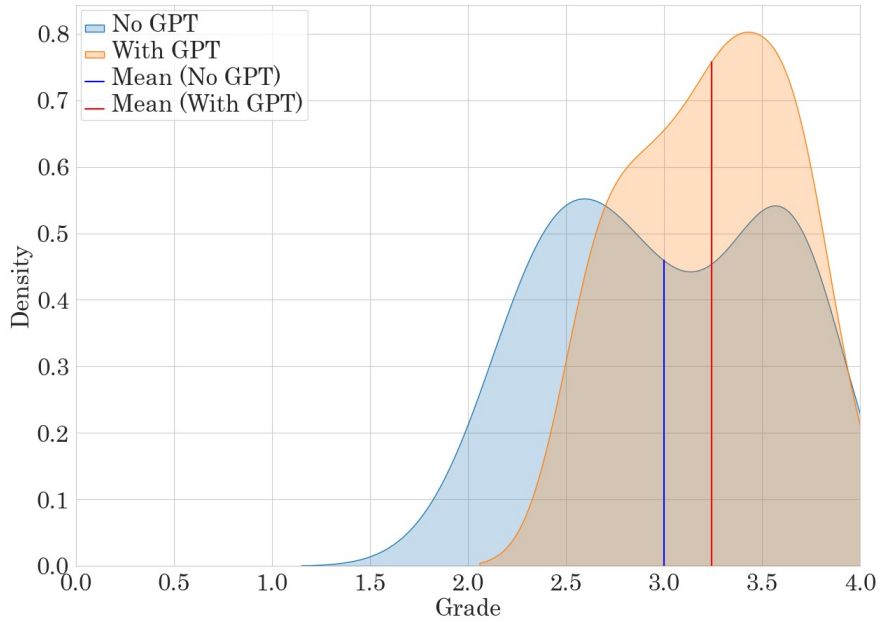


Figure 3: Quality Distributions with and Without AI—Employee Handbook

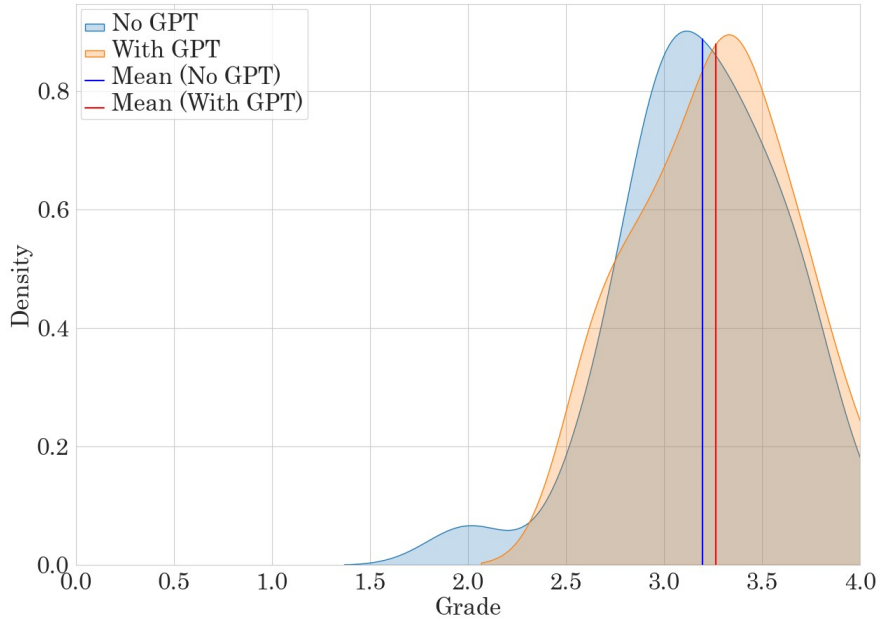


Figure 4: Quality Distributions with and Without AI—Client Memo

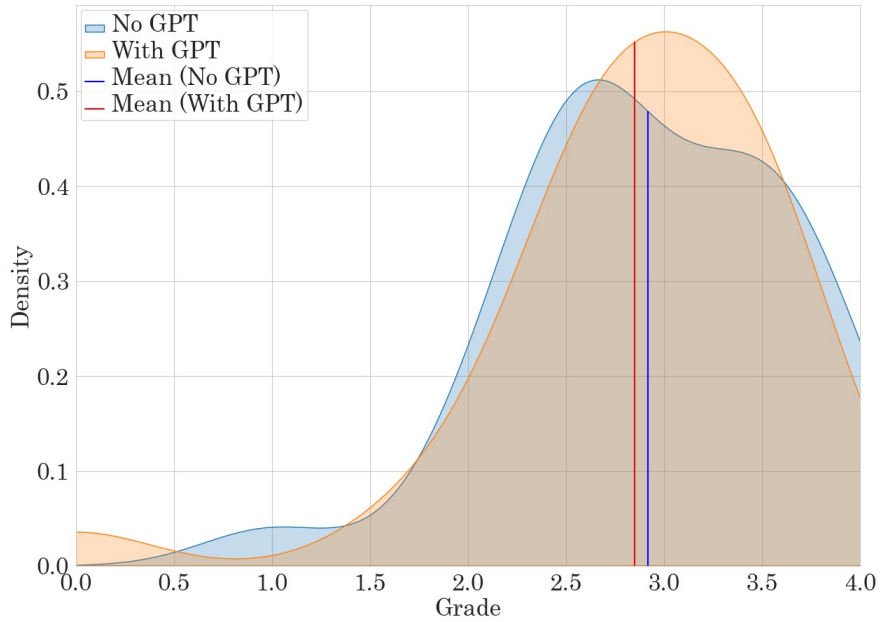


Table __ below depicts the effect of access to GPT on the amount of time taken on each task. These results are more decisive, showing large and consistent decreases in the amount of time taken on each task. Interestingly, the largest gain in speed (in percentage terms) occurs in the task for which GPT-4 was the most useful in terms of grade improvement (contract drafting), and the smallest gain in speed (again in percentage terms) occurs in the task for which GPT-4 was the least useful (client memo).

Table 2: Effect of GPT-4 on Time Taken (Minutes)

	No GPT (Std. Dev.)	With GPT (Std. Dev.)	Difference (95% CI)	% **Difference** (95% CI)
Complaint Drafting	160.69	122.00	-38.77	24.1%
	(72.38)	(66.80)	(-64.00, -13.36)	
Contract Drafting	69.72	47.59	-22.40	32.1%
	(32.00)	(31.09)	(-33.71, -10.91)	
EE Handbook	37.24	29.41	-7.84	21.1%
	(9.55)	(13.42)	(-12.03, -3.74)	
Client Memo	244.41	215.69	-28.75	11.8%
	(58.03)	(72.96)	(-52.59, -5.05)	

Figures __ through __ show the distributions of the amount of time that participants took on each task. Figures __ through __ in the Appendix show bootstraps for the differences in means between groups, showing that the decrease in the time participants took on every task is statistically significant at the 95% level.

Figure 5: Time Distributions with and Without AI—Complaint Drafting

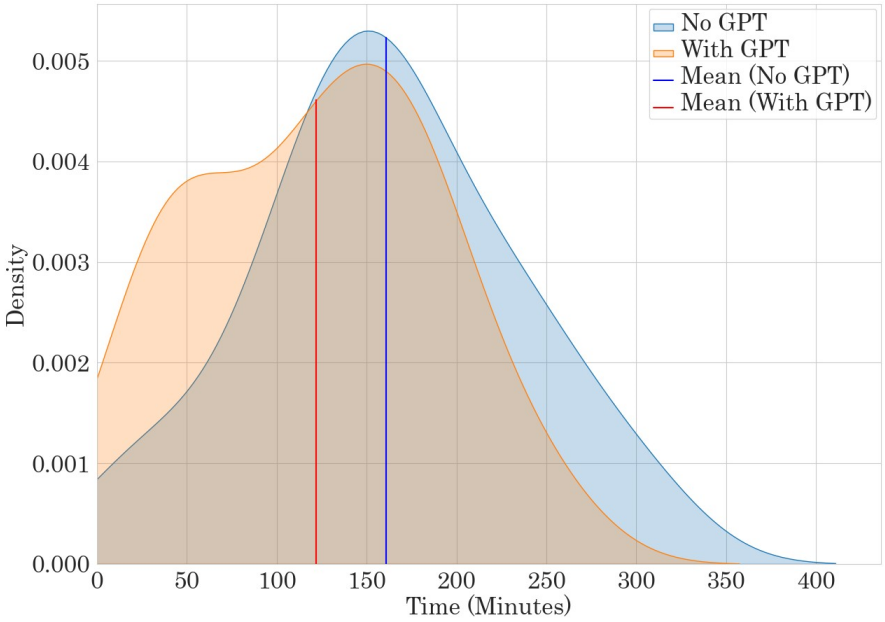


Figure 6: Time Distributions with and Without AI—Contract Drafting

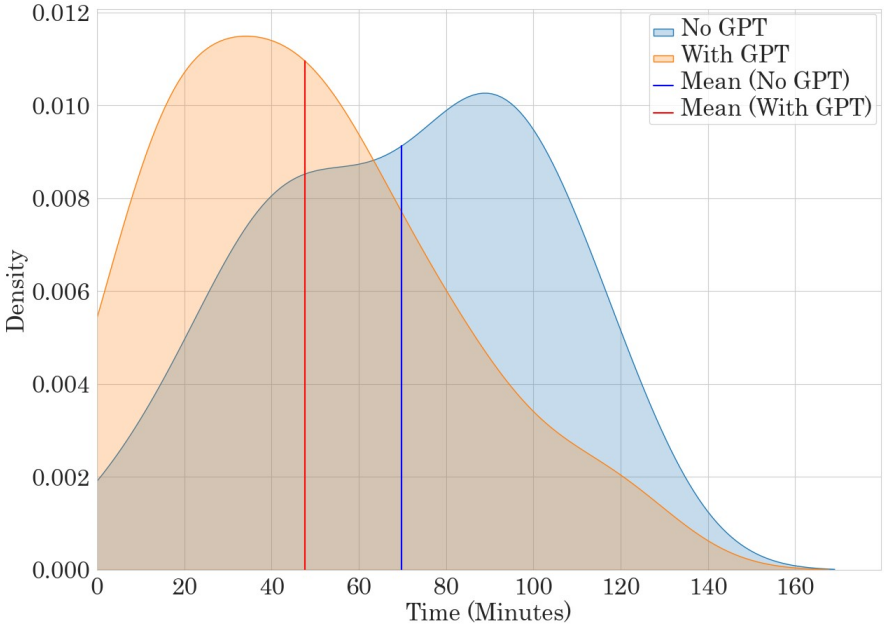


Figure 7: Time Distributions with and Without AI—Employee Handbook

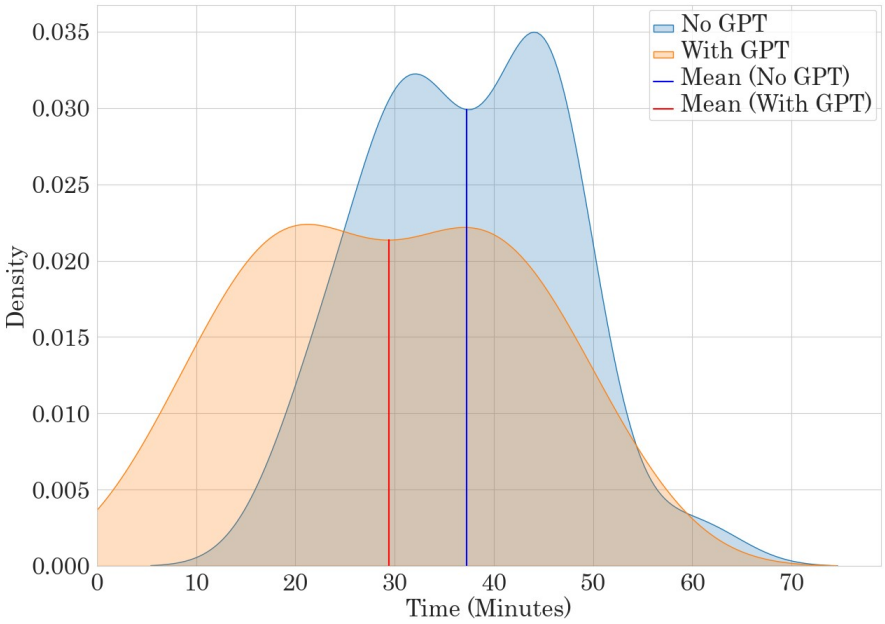
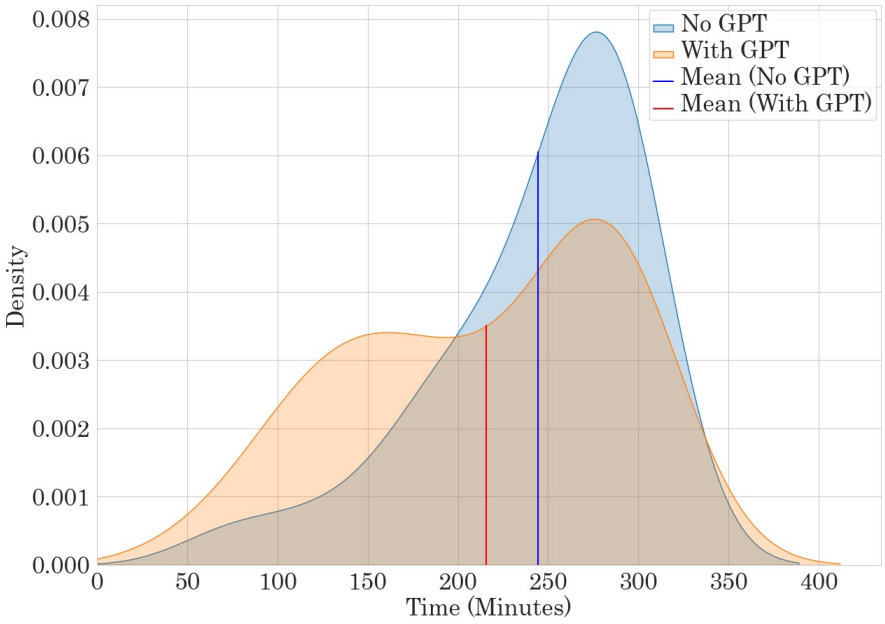


Figure 8: Time Distributions with and Without AI—Client Memo



In addition to raw results comparing the groups that did and did not have access to GPT-4, we can also evaluate how the effect of AI

assistance on performance and time taken varied *within* each group. Namely, we can test whether the boost provided by GPT-4 was larger for participants who performed better without access to GPT-4. To conduct this comparison, we graph performance at one task against performance at another task. Recall that each participant completed two tasks with the aid of GPT-4 and two tasks without access to the AI. We should expect that performance at one legal task should somewhat predict performance at any other legal task. Thus we can first take each participant's grade at one task they conducted without GPT-4 (graphed on the x-axis) and compare that against their performance at the other task without GPT-4 (graphed on the y-axis). This creates a baseline that we can use as a control to establish how replicable performance is in the absence of access to AI, shown as the blue line in Figures __ through __ below. Conceptually, if performance is perfectly correlated between tasks, this line should be a 45-degree angle where $x = y$. The graphs are separated based on which task was used as Task 2.

We can then take the two tasks that each participant completed without access to AI and use them to graph another line, showing how their performance on a task without GPT-4 (on the x-axis) predicts performance with access to GPT-4 (on the y-axis). This is the red line in the figures below.[53] For each of the following Figures, Task 2 is held constant for each graph, while Task 1 includes participants' performance on the other relevant tasks. Thus, given each participant's actual grade on a different task (located on the x-axis), the corresponding point on the blue line on the y-axis is their expected grade on Task 2 without GPT-4's assistance, and the corresponding point on the red line on the y-axis is their expected grade *with* GPT-4's assistance. This means, for instance, that if the red line is consistently higher than the blue line, the expected benefit from using GPT-4 is positive regardless of baseline skill level.

Most importantly, the relative slopes of the red and blue lines tell us whether or not GPT-4 acts as an equalizing force. If AI assistance flattens the distribution of performance, the red line will be flatter than the blue line; if AI has no effect on the distribution of performance, the red line should run parallel to the blue control line. The *difference* in the slopes of the blue and red lines measures the extent to which access to GPT-4 flattens performance.

[53] The range of the treatment and control lines on the x-axis differ for some of the graphs, because there are some tasks on which

Figure 9: Task 1 vs. Task 2 Grades—Complaint Drafting

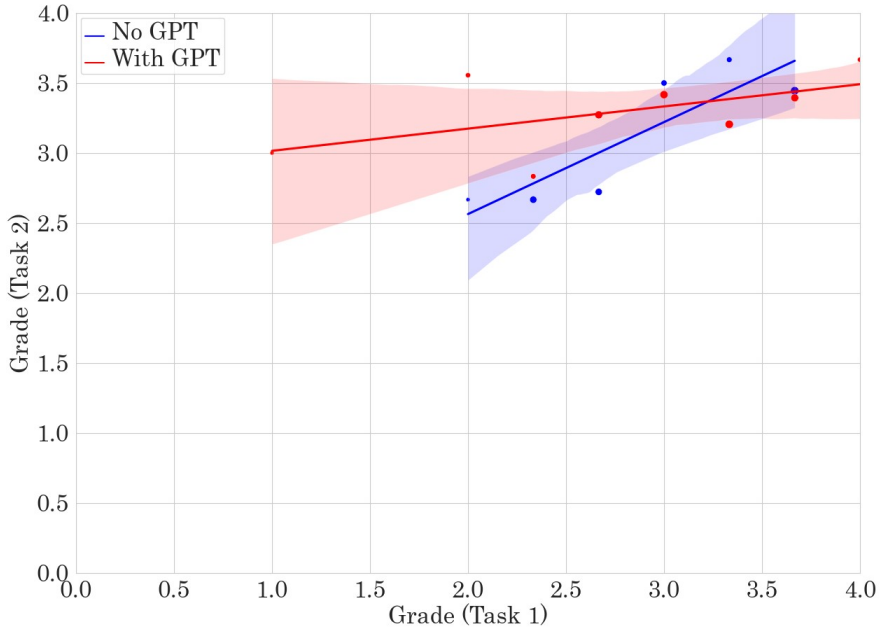


Figure 10: Task 1 vs. Task 2 Grades—Contract Drafting

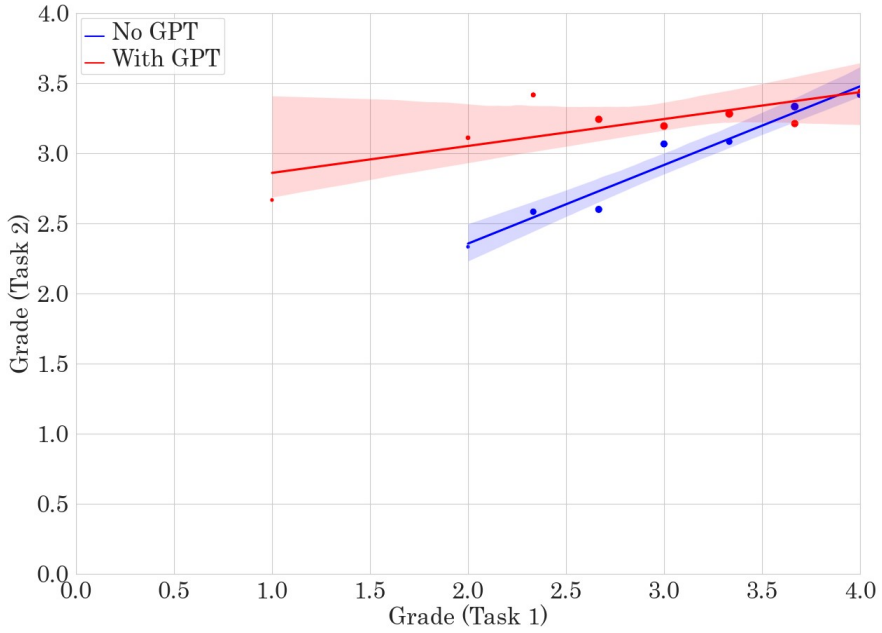


Figure 11: Task 1 vs. Task 2 Grades—Employee Handbook

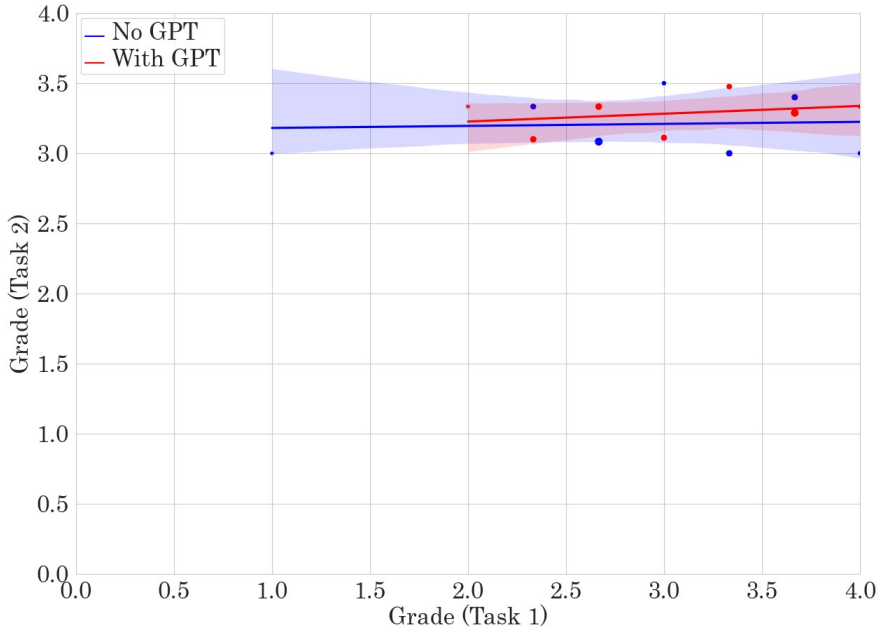
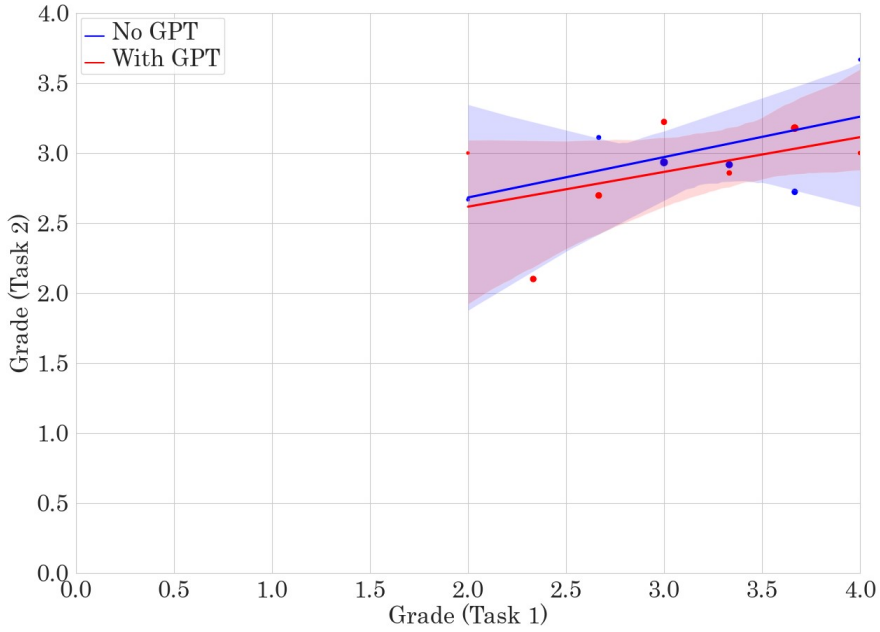


Figure 12: Task 1 vs. Task 2 Grades—Client Memo



As the Figures show, where GPT-4 assistance provided some benefit, that benefit was unequally distributed. On the tasks where GPT-4 was most useful (the contract drafting and complaint drafting tasks)

the slope of the line with access to GPT is substantially flatter than the line without, indicating that GPT-4 provides a greater boost to low performers than high performers. On the tasks where GPT-4 had near zero effect on performance (the client memo and EE handbook tasks) the slopes of the treatment and control lines are almost identical, indicating that access to GPT-4 had roughly the same impact regardless of baseline performance—that is, no impact.

In sum, where assistance from GPT-4 is beneficial at all, it seems to benefit the worst performers the most, providing little or no benefit to top performers. Table __ below confirms that the differences in slopes are large and statistically significant at the 95% level.

Table 3: Slope of GPT-4 on Performance (Grade)

	No GPT (95% CI)	With GPT (95% CI)	Difference (95% CI)
Complaint Drafting	0.66	0.16	0.50
	(0.35, 0.95)	(0.00, 0.28)	(0.20, 0.84)
Contract Drafting	0.56	0.19	0.37
	(0.33, 0.80)	(-0.06, 0.20)	(0.22, 0.74)
Employee Handbook	0.01	0.06	-0.05
	(-0.21, 0.19)	(-0.03, 0.21)	(-0.33, 0.13)
Client Memo	0.29	0.25	0.01
	(-0.64, 0.48)	(0.25, 0.75)	(-1.16, 0.06)

We can conduct the same sort of analysis for the effect of AI assistance on the amount of time taken to complete each task, shown in Figures __ through __ below. Because each task took a different amount of time on average, we scaled the raw minutes spent by dividing them by the mean minutes spent per task (whether with GPT-4 or without), in order to be able to aggregate different tasks into Task 1 and to make the slopes directly comparable. Although access to GPT-4 consistently decreased the time taken on each task (the red lines are consistently below the blue lines), they are generally parallel, indicating no leveling effect on the amount of time taken depending on the baseline amount of time taken. The one exception is contract drafting, where there is a difference in slopes, although it is not statistically significant at the 95% level.

Figure 13: Task 1 vs. Task 2 Time—Complaint Drafting

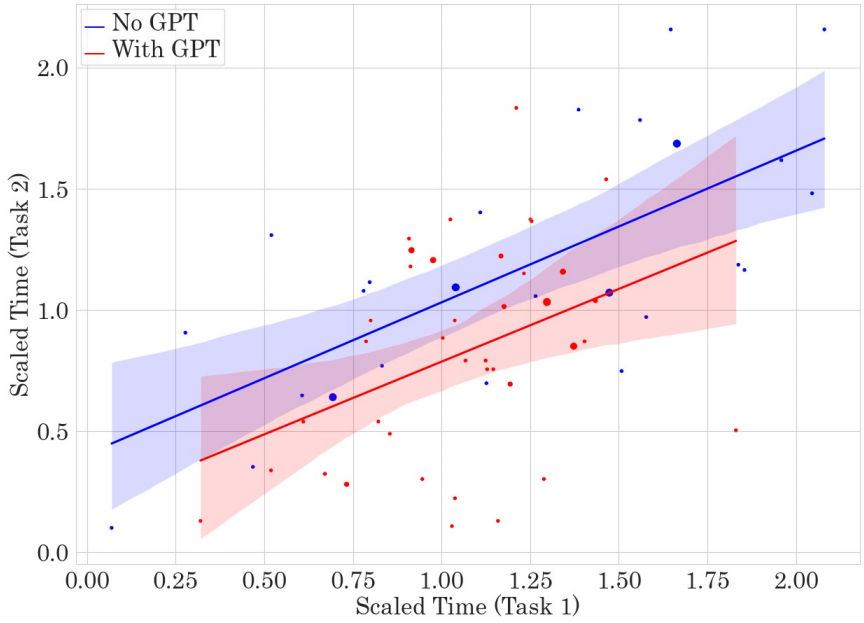


Figure 14: Task 1 vs. Task 2 Time—Contract Drafting

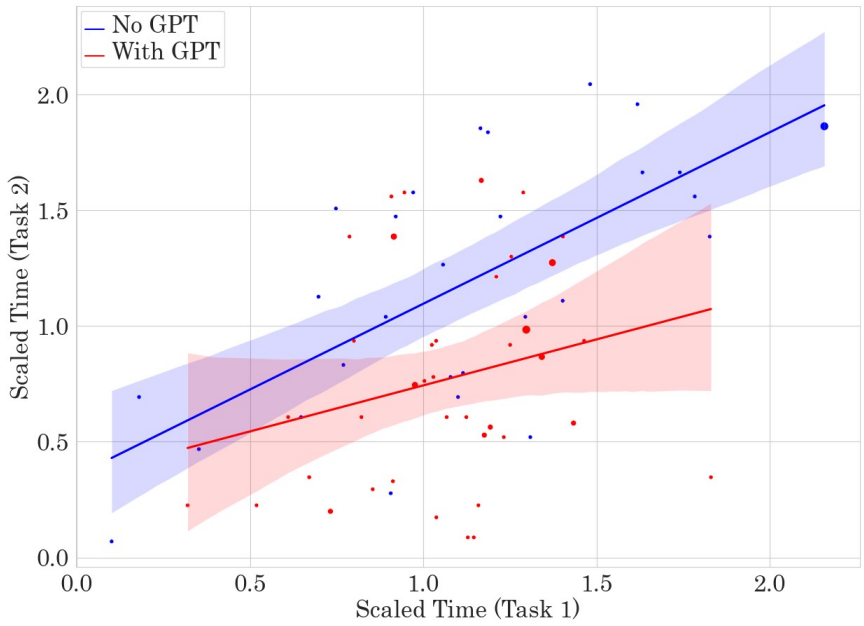


Figure 15: Task 1 vs. Task 2 Time—Employee Handbook

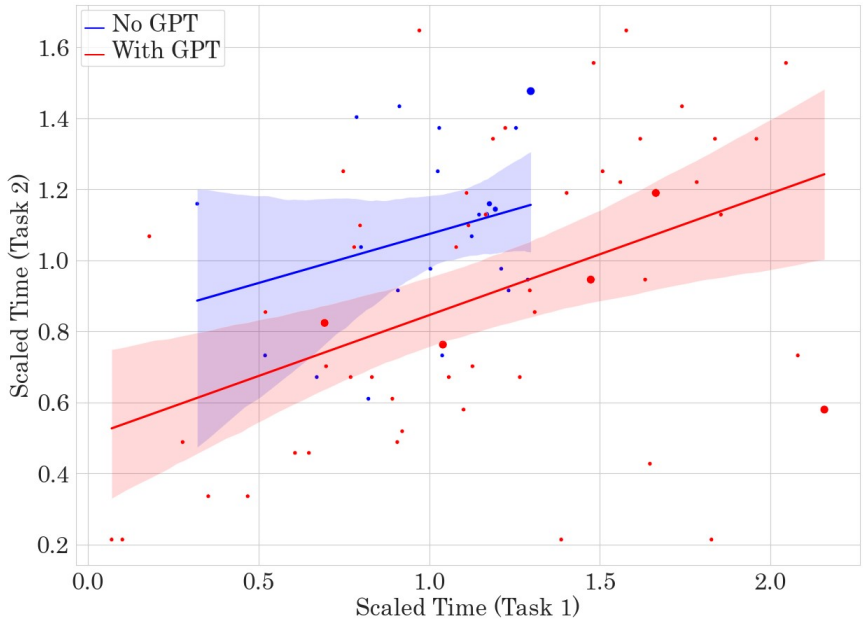


Figure 16: Task 1 vs. Task 2 Time—Client Memo

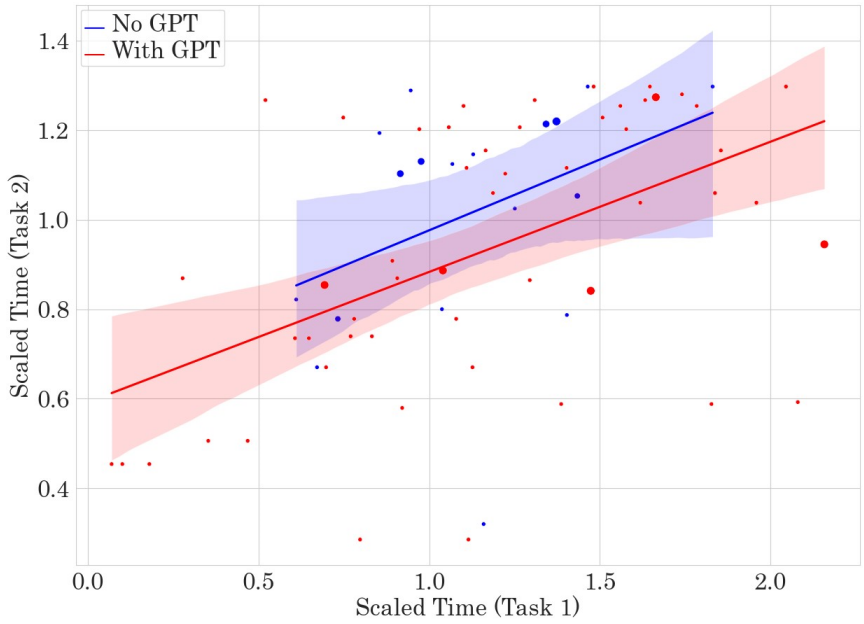


Table __ reflects these results.

Table 4: Slope of GPT-4 on Performance (Grade)

	No GPT (95% CI)	With GPT (95% CI)	Difference (95% CI)
Complaint Drafting	0.63	0.60	0.03
	(0.39, 0.90)	(0.26, 0.88)	(-0.32, 0.48)
Contract Drafting	0.74	0.40	0.34
	(0.52, 0.96)	(0.12, 0.75)	(-0.08, 0.68)
Employee Handbook	0.28	0.34	-0.06
	(0.05, 1.03)	(0.16, 0.45)	(-0.30, 0.71)
Client Memo	0.32	0.29	0.03
	(0.07, 0.58)	(0.18, 0.38)	(-0.22, 0.31)

Finally, we surveyed study participants on their perceptions of GPT-4 based on the assignments. The specific survey questions were:

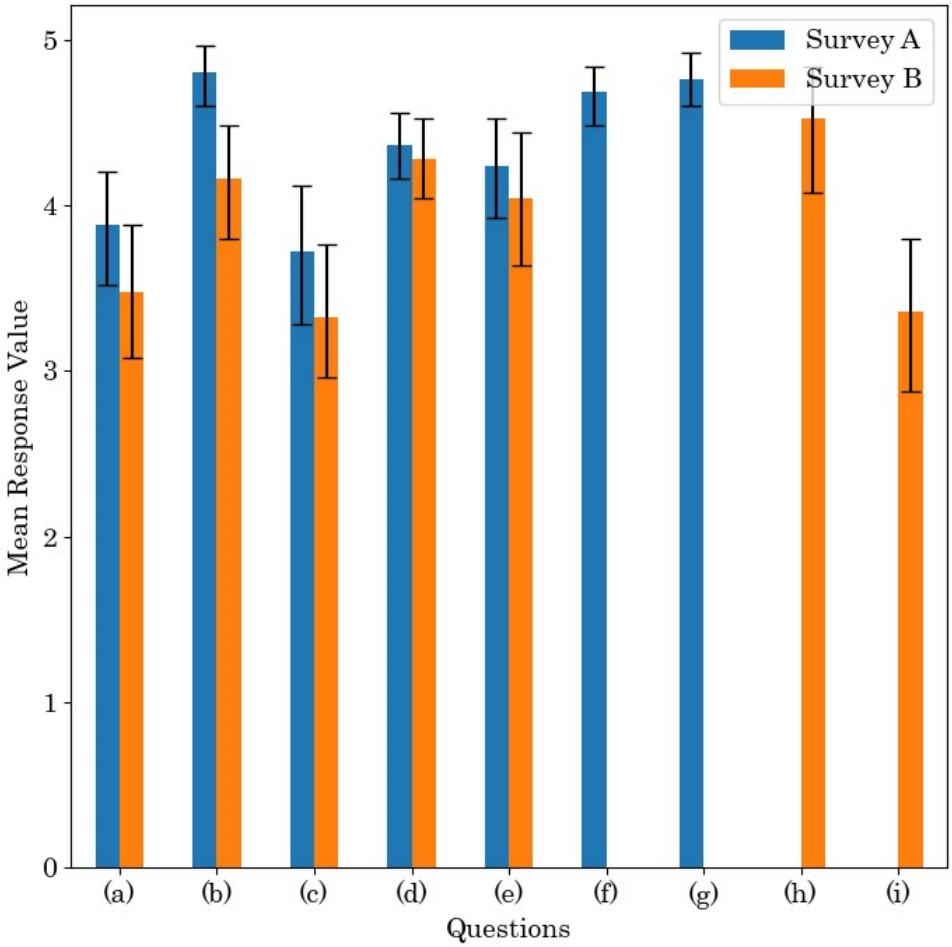
Survey Questions

a) *For the assignments on which you had access to GPT-4, to what extent did this access impact the quality of the work that you completed for these assignments?*
b) *For the assignments on which you had access to GPT-4, to what extent did this access impact the speed with which you could complete the assignments?*
c) *For the assignments on which you had access to GPT-4, to what extent did this access impact the personal satisfaction that you experienced in completing these assignments?*
d) *To what extent did you find that your ability to use GPT-4 effectively for legal drafting improved over the course of the experiment?*
e) *How did your experience in this experiment impact the extent to which you anticipate using tools like GPT-4 for legal work in the future?*
f) *To what extent did you find access to GPT-4 to be helpful for the complaint drafting assignment specifically?*
g) *To what extent did you find access to GPT-4 to be helpful for the contract drafting assignment specifically?*
h) *To what extent did you find access to GPT-4 to be helpful for the Employee Handbook drafting assignment specifically?*
i) *To what extent did you find access to GPT-4 to be helpful for the Legal Memo drafting assignment specifically?*

Participants responded to these questions using a Likert scale with 5 values: substantially no, somewhat no, neither yes nor no, somewhat yes, and substantially yes (with appropriate modification based on the wording of the specific question). Figure __ shows the results

of this survey, coding the Likert responses on a 1-5 scale, with 95% confidence intervals indicated by brackets[54]:

Figure 17: Survey Results



These results are interesting along several different dimensions. First, recall that Group A and Group B had access to GPT-4 on different assignments, and that Group A used GPT-4 for the tasks on which it was generally most effective (contract drafting and complaint drafting). Consistent with those assignments, Group A reported on average that GPT-4 had a larger effect both on the quality and speed of their work. Participants in Group A also reported a larger boost to personal satisfaction when provided access to GPT-4. Both groups reported that

[54] The 95% confidence intervals were generated through bootstrapping with 10,000 iterations.

their ability to use GPT-4 improved over the course of the assignments and that participating in the study made them more likely to use GPT-4 for future work. Finally, respondents accurately perceived how useful GPT-4 was for specific tasks. In fact, the ordinal ranking of the impact of AI assistance on task performance exactly corresponds with the ranking of how useful participants perceived AI to be on each task, with contract drafting ranked the highest and the client memo ranked the lowest.

IV. IMPLICATIONS

A. Implications for the Future of Legal Services

Our findings show that providing law students with general purpose and widely available generative AI tools like GPT-4 and a limited amount of training can substantially improve the efficiency with which they complete a broad array of legal tasks without adversely affecting (or even slightly improving) the quality of that work product. Moreover, they suggest that young lawyers provided with access to AI to facilitate their work accurately appreciate these benefits of AI, find that access to AI tends to enhance their work satisfaction, and generally become more enthusiastic about using AI to facilitate their work as they gain experience doing so.

Standing alone, these results suggest that generative AI will almost certainly become a vital tool for many lawyers in the near future, comparable to more familiar legal-tech tools like Westlaw, Lexis and e-discovery software.[55] Indeed, this trend has already begun, with some lawyers and law firms proactively embracing generative AI.[56] For less proactive lawyers and firms, our results suggest that the embrace of AI will likely be driven by competitive dynamics, as legal services providers that embrace AI can charge lower rates or deliver more, or higher quality, results than their less-forward-looking competitors.

The implications of our results become substantially more striking, however, when they are considered in light of the current pace

---

[55] See Part I, supra.

[56] See, e.g., Kate Beioley & Cristina Criddle, *Allen & Overy Introduces AI Chatbot to Lawyers in Search of Efficiencies*, Fin. Times (Feb. 14, 2023), https://www.ft.com/content/baf68476-5b7e-4078-9b3e-ddfce710a6e2 [https://perma.cc/66NV-N6UD; Emily Hinkley, *Mishcon de Reya Is Hiring an 'Engineer' to Explore How Its Lawyers Can Use ChatGPT*, Legal Cheek (Feb. 16, 2023, 8:35:00 AM), https://www.legalcheek.com/2023/02/mishcon-de-reya-is-hiring-an-engineer-to-explore-how-its-lawyers-can-use-chatgpt [https://perma.cc/G2HE-H8CY].

of innovation in AI generally, and legal AI in particular. This is because our results are likely to substantially *understate* the future potential of AI to aid in the provision of legal services in at least three different respects.

First, and most importantly, whereas our results focused on the impact of GPT-4 on the provision of legal services, numerous more specialized generative AI tools for lawyers are already widely available, and many more are under development.[57] Currently available tools offer lawyers vastly superior capabilities than the general-purpose AIs like GPT-4 that we used in our experiment. They accomplish this predominantly by marrying generative AIs like GPT-4 with intelligent prompt-engineering and Retrieval Augmented Generation (RAG). The former technique bakes into legal tech platforms prompting strategies that are tested and customized to produce useful results for specific types of legal tasks.[58] RAG, the latter approach, allows generative AIs to retrieve relevant content from large legal databases and to use this material to inform its responses.[59] Combined, these two techniques

---

[57] For instance, the firm Casetext recently launched a product known as CoCounsel, which "does document review, legal research memos, deposition preparation, and contract analysis in minutes—with results you can trust." Casetext, https://casetext.com [https://perma.cc/5SDR-PG3S]. Within months of Counsel's launch, the legal tech giant Thompson-Reuters purchased Casetext for $650 Million. See, e.g., Thomson Reuters to Acquire Legal AI Firm Casetext for $650 Million, REUTERS (June 27, 2023), https://www.reuters.com/markets/deals/thomson-reuters-acquire-legal-techprovider-casetext-650-mln-2023-06-27.

[58] For general literature on prompt engineering, see Dils, *How to Use ChatGPT: Advanced Prompt Engineering*, WGMI Media (July 20, 2023), https://wgmimedia.com/how-to-use-chatgpt-advanced-prompt-engineering [https://perma.cc/3ZER-KRCW]; Tyler Cowen & Alexander T. Tabarrok, How to Learn and Teach Economics with Large Language Models, Including GPT (March 17, 2023) (unpublished manuscript) (on file with authors). For prompt-engineering advice that is specific to the legal setting, see Daniel Schwarcz & Jonathon Choi, *AI Tools for Lawyers: A Practical Guide*, 108 Minn. Law Review Online (forthcoming, 2023).

[59] See generally Patrick Lewis et al, Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks 33 Advances in Neural Information Processing Systems (NeurIPS 2020). For discussion of how tools like Casetext use RAG, see https://casetext.com/blog/prompt-engineering-best-ai-output/ ("By connecting GPT-4 to a database of

substantially reduce hallucinations and improve the quality of AI-generated output.

A second way in which our results understate the potential of AI to improve the efficiency of legal services is that our study participants had limited experience using this technology. In total, participants in our study received a couple hours of online training before attempting to use this technology to craft answers to two of the four assignments they completed while participating in the study.[60] Not surprisingly, participants did not believe that this training fully equipped them to use generative AI effectively and efficiently, as illustrated by their survey results indicating that their ability to use AI improved over the course of the experiment.[61] By contrast, as lawyers and law students use generative AI in their practice, they will naturally tend to become more adept at using it effectively and efficiently.[62]

A third and final reason that our results understate the transformative potential of AI in legal services is that the capabilities of generative AI—which we measured in the summer of 2023—are continuing to rapidly accelerate.[63] To illustrate, GPT-4, which OpenAI released in March 2023, is significantly better at legal analysis than GPT-3.5, the model that open AI released only several months earlier in late 2022.[64] LLMs are almost certain to improve in the coming years due

---

reliable legal sources, we're able to ground its output in real-world knowledge rather than leaving it to rely only on its own memory.").

[60] See Part II, supra.

[61] See Part III, supra.

[62] See Schwarcz & Choi, supra note X ("The quickest route to proficiency with LLMs is the same route to Carnegie Hall: practice, practice, practice.").

[63]

[64] See, e.g., Daniel Martin Katz et al., *GPT-4 Passes the Bar Exam* (Mar. 15, 2023) (unpublished manuscript) (on file with authors); Jonathon Choi & Daniel Schwarcz, AI Assistance in Legal Analysis: An Empirical Study, Journal of Legal Education (forthcoming, 2024); Andrew Blair-Stanek et al., GPT-4's Law School Grades: Con Law C, Crim C-, Law & Econ C, Partnership Tax B, Property B-, Tax B (May 24, 2023) (unpublished manuscript) (on file with authors). For examples demonstrating that GPT-4 outperforms ChatGPT in other fields, see Chung Kwan, What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature, 13 EDUC. SCI. 410 (2023); David A. Wood et al., The ChatGPT Artificial Intelligence Chatbot: How Well Does It Answer Accounting Assessment Questions?, 2023 ISSUES IN ACCOUNTING EDUC. 1; Harsha Nori et al., Capabilities of GPT-4 on Medical Challenge Problems (Apr. 12, 2023) (unpublished manuscript)

to increases in model size and complexity and continuing innovation in the underlying AI architecture.

Although these three considerations lead us to conclude that our results significantly understate the transformative potential of AI in the legal space, there are several considerations pointing in the other direction. Most notably, participants in our experiment self-selected from the general population of law students at the University of Minnesota in response to an email soliciting interest in an experiment on AI and the law. These students thus were likely to have been particularly enthusiastic about the potential of generative AI in law, and thus potentially more likely than the population of law students as a whole to benefit from using AI. It is also possible that these participants disproportionately had some prior exposure to using generative AI to complete legal tasks.

Not only do our results suggest that generative AI will produce significant efficiencies across a broad range of legal services, but they also imply that these efficiencies will be distributed highly unevenly across practice areas, task types, and lawyer skill levels. This conclusion follows from two of our bottom-line findings. First, the boost in quality experienced by participants was higher for participants with a lower baseline skill set than for those with a higher baseline skill set.[65] This result is consistent both with some of our own prior work in the legal arena, as well as with a number of high-profile studies examining how access to AI impacts the quality of work product outside of the legal arena, for workers such as professional writers, customer service agents, and medical professionals.[66] If AI assistance most benefits the least-

(on file with authors); Alejandro Lopez-Lira & Yuehua Tang, Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models (May 12, 2023) (unpublished manuscript) (on file with authors).

[65] See Part III, supra.

[66] See Jonathon Choi & Daniel Schwarcz, AI Assistance in Legal Analysis: An Empirical Study, Journal of Legal Education (forthcoming, 2024) (reporting "significant variation in how useful AI assistance was to students depending on their baseline performance," with "worst-performing students benefited enormously from AI, with gains of approximately 45 percentile points," while " the best-performing students received worse grades when given access to AI, experiencing declines of approximately 20 percentile points"). For literature outside of the legal setting finding uneven quality gains from access to AI based on the baseline skill of workers, see Shakked Noy & Whitney Zhang, *Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence*, 381 SCIENCE 187, 187 (2023) (finding that giving college-

skilled lawyers, then it follows that AI is likely to serve as an equalizing force in a notoriously unequal profession.

Here too, our results are likely to understate the extent to which access to generative AI will have variable effects for different subsets of lawyers across different practice areas. This is because participants in our study represented a very narrow and relatively homogenous subset of the legal profession: current or just-graduated students at the University of Minnesota Law School in the summer of 2023. All such students, of course, gained admission to the law school, meaning that they almost uniformly performed exceptionally well both with respect to their college grades and the LSAT examination. The range of baseline skillsets possessed by legal professionals in general varies much more dramatically than was the case for our study participants. This point is mitigated by the fact that participants in our study were disproportionately inexperienced relative to average legal professionals, but only moderately so given that our focus was on relatively simple legal tasks that would tend to be assigned to junior attorneys.

Second, we found that AI enhanced the quality of participants' work product significantly more for some tasks (contract drafting in particular) than others, where it had limited or no effect on quality (legal memo and employee handbook). This result is also consistent with some of our own prior research, which found that providing humans with AI produced significant gains in accuracy with respect to simple multiple-choice questions, limited quality gains for straight-forward legal essays, and no average gains in quality with respect to student answers to complex and advanced legal essay questions.[67]

---

educated professionals access to AI improved the performance of less skills workers more than high skilled workers); Erik Brynjolfsson, Danielle Li & Lindsey R. Raymond, *Generative AI at Work* (Nat'l Bureau of Econ. Rsch., Working Paper No. 31161, 2023) (finding that giving customer service agents access to AI improved the capabilities of less skilled agents more than highly skilled agents). See also Fabrizio Dell'Acqua, Falling Asleep at the Wheel: Human/AI Collaboration in a Field Experiment on HR 1 (Dec. 2, 2021) (unpublished manuscript) (on file with authors) (finding that giving professional recruiters access to high quality AI harmed humans' ability to assess job applications relative to giving them access to less high quality AI toolss

[67] See Choi & Schwarcz, AI Assistance in Legal Analysis, supra (finding that AI produced significant gains in quality when provided to undergraduates answering basic law school style questions, minimal average gains in quality with respect to undergraduate answers to straight-forward legal essays, and less still with respect to upper level law students' answers to more complex legal questions).

Once again, the uneven average impact of AI on quality across task types is likely to be understated by our results. That is because all four of the legal tasks we selected for the study necessarily shared certain features given our experimental design: they required a written work product, necessitating little if any independent research, that could be completed in between 1 to 5 hours of time, and that were reasonably appropriate for law students. These constraints, of course, do not apply to the immense range of tasks that real lawyers may need to complete. The features of some lawyer tasks—such as negotiating complex deal terms or crafting high-stakes legal briefs—almost certainly make them less amenable to assistance from AI. Meanwhile, many other legal tasks are likely to be much more dramatically impacted by the availability of AI than those that we focused on in our experimental setting. One important example involves the simple act of summarizing large and complex documents, such as deposition transcripts. General purpose AIs are particularly adept at summarizing complex and dense material, and specialized AI tools like CoCounsel use basic prompt engineering strategies to improve the reliability and verifiability of these efforts.[68] Anecdotal reports from lawyers indicate that these tools can perform certain summarization tasks that would ordinarily take a young associate hours in a matter of minutes, while producing more reliable output.

Finally, participants in our study accurately assessed how useful GPT-4 was at each task and reported increased satisfaction when using tasks with access to GPT-4. This suggests that lawyers will be able to apply AI effectively when it is most useful and that AI could improve lawyer wellbeing by taking on relatively tedious work.

In sum, when considered in light of current trends in the development of generative AI as well as prior research, our results suggest that the practice of law is on the precipice of significant—and potentially foundational—change and transformation. This change will, however, occur unevenly across legal domains and practice areas.

Importantly, these predictions concern only the first-order impacts of generative AI on the legal profession: legal technologies built on generative AI will become a vital and potentially transformative tool for a broad range of lawyers. The higher-order impacts of this reality are, of course, much harder to predict. Will demand for legal services increase or decrease? Will firms alter the range of legal services that they send to outside counsel relative to the tasks that they perform in house? Will lawyer pay become higher, lower, or more uneven? And what impact will all of the above have on the demand and supply of lawyers and law students? Our empirical results offer limited guidance on these

[68] See CoCounsel features (summarizing tool)

questions, other than to suggest that the assumption that the future will resemble the past is likely tenuous, at best.

B. Normative Implications

Lawyers, judges, clients, law schools, and law students will all need to adjust over the coming years as tools that incorporate generative AI become a reality of legal practice. Of course, both the pace and the character of this innovations remain deeply uncertain. But our results provide some helpful context regarding how individual actors within the legal system can and should adapt to this transformation in the near term.

1. Law Schools and Law Students

The training that law schools provide to their students has in many ways remained unchanged for more than a century. This is particularly true when it comes to first-year law students, who have long studied the same mandatory curriculum, which is typically taught to them through some form of Socratic instruction.[69] Although recent decades have seen important adaptations to this approach—from more inclusive Socratic questioning,[70] to an increased focus on statutory interpretation,[71] to increased opportunities for formative feedback[72]—none of these changes have fundamentally altered the character of legal education, particularly in the first-year of law school.

In our view, this consistency in basic legal pedagogy properly reflects a consistency in the basic features of effective legal reasoning.[73] Not even technological change as significant as generative AI is likely to

---

[69] See L. Danielle Tully, *What Law Schools Should Leave behind*, 2022 Utah L. Rev. 837 (2022); Rachel Gurvich , L. Danielle Tully , Laura A. Webb , Alexa Z. Chew, Jane E. Cross & Joy Kanwar, *Reimagining Langdell's Legacy: Puncturing the Equilibrium in Law School Pedagogy*, 101 N.C. L. Rev. F. 118 (2022-2023).

[70] Jamie R. Abrams, Legal Education's Curricular Tipping Point Toward Inclusive Socratic Teaching, 49 Hofstra L. Rev. 897, 898 (2021).

[71] Abbe R. Gluck, The Ripple Effect of "Leg-Reg" on the Study of Legislation & Administrative Law in the Law School Curriculum, 65 J. Legal Educ. 121 (2015).

[72] See Daniel Schwarcz & Dion Farganis, *The Impact of Individualized Feedback on Law Student Performance,* 67 J. Legal Educ. 139 (2017).

[73] See, e.g., Cass R. Sunstein, On Analogical Reasoning, 106 Harvard Law Review 741 (1993).

alter this reality any time soon. To the contrary, effectively using AI to craft legal arguments requires many of the same basic legal and analytical skills as other forms of lawyering, including a capacity to question initial answers, confirm the accuracy of arguments and sources, organize issues clearly, and assess the strength of alternative arguments.[74]

For these reasons, we believe law schools should ban or substantially limit the use of generative AI in conventional first-year law school classes. Because generative AI does not impact the nature of legal reasoning, it should not alter the way that such reasoning is taught by instructors or demonstrated by students, particularly introductory law students. In many ways, this pedagogical approach should be familiar: for instance, introductory math students are universally taught to add, subtract, multiply and divide without the aid of calculators, as mastering these basic skills is essential for most forms of higher math.

However, our results suggest that accomplishing this goal requires law schools to proactively limit access to generative AI during student assessments for formative classes. That is because they demonstrate that generative AI can not only empower law students to craft legal work product significantly more quickly (a skill that is typically rewarded on timed law school exams), but also that it can disproportionately improve the quality of that work product for less skilled students. Our prior work has demonstrated that this is true not only for the practical legal tasks that we focused on in this experiment, but also for a range of different types of law school exams.[75] Thus there is a risk that students will use AI as a crutch rather than developing crucial lawyering skills early in their careers. In addition, AI assistance will tend to compress the distribution of grades in traditional law school exams and make it more difficult for professors to provide individualized feedback.

Given these realities, we believe that law professors who intend to limit access to AI in student assessment must technologically limit students' abilities to access generative AI tools. Relying instead on honor codes backed by the threat of enforcement is simply impractical given the current power of widely accessible generative AI tools. This is especially so because there are currently no reliable tools available for identifying content produced by generative AI, meaning that law schools and professors cannot reliably detect cheating.[76] All of this means that cheating among a non-trivial number of students is inevitable when

[74] See Schwarcz & Choi, AI Tools for Lawyers: A practical guide, supra.

[75] See AI Assistance in Legal Analysis.

[76] See, e.g., https://arxiv.org/pdf/2305.03807.pdf

instructors rely only on an honor code to prevent student use of generative AI. Over time, we fear that such cheating among a handful of students would spread as students who were initially inclined to follow the rules begin to feel like "suckers" for doing so.

While our results lead us to conclude that law schools should take affirmative measures to restrict student access to generative AI tools in some classes, we also believe that they suggest that law schools should simultaneously develop upper-level classes that explicitly train students on how to use generative AI tools effectively. This conclusion is buttressed by our survey results indicating that participants reported that their ability to use AI effectively increased markedly over the course of the experiment, that participating in the experiment increased their interest in using AI in their future work, and that using this tool also increased their personal satisfaction.[77] It is also supported by the differential impact of AI on quality across the different task types; whereas students interested in some practice areas may rightly believe that it would not be a good use of their law school credits to take a class that focuses significant attention on using generative AI, other students may rightly reach the opposite conclusion depending on their career aspirations and interests.

The quantity and scope of these classes should of course vary by school and context, though law schools with students who are more interested in or likely to provide legal services to individuals or cost-sensitive clients should be particularly aggressive in developing these course offerings. So too should law schools that focus on producing "practice-ready" attorneys who are less likely to receive extensive on-the-job training early in their career. Although the supply of instructors who are comfortable teaching classes on how to use generative AI in the law may be limited at first, we suspect that this pool of potential instructors will grow as does the use of generative AI in practice. Moreover, a virtue of generative AI tools is that those with significant legal expertise may be better positioned than they initially believe to learn how to use these tools effectively along with their students.

2. Lawyers and Law Firms

Our results strongly suggest that virtually all lawyers and law firms should be proactively exploring how best to incorporate generative AI tools into their practice. Of course, many law firms are doing just that. For instance, in March of 2023, the global law firm DLA Piper announced that it would incorporate CoCounsel, one of the leading generative AI

[77] See Part III, supra.

tools for lawyers, into its practice.[78] Numerous other large law firms have also embraced this tool in recent months, though many have been reluctant to publicly acknowledge this.[79] Other large global law firms—including Allen & Overy—have incorporated a competing generative AI tool, Harvey, into their practice.[80] Still other firms have taken a different approach, hiring their own AI experts to develop proprietary and firm-specific generative AIs that are not available to competitors.[81]

Although this trend is already evident in big-law, at least some smaller law firms and solo-practitioners have also begun exploring how to incorporate generative AI into their work, with mixed results. The most notorious such example involved a lawyer who relied on ChatGPT to author a brief without double-checking the resulting output. The generative AI proceeded to hallucinate the existence of several cases, and then to insist on questioning from the lawyer that these cases were real. Not surprisingly, the unwitting lawyer was publicly excoriated by the judge in a hearing that was reported on widely by the media and that drew widespread attention from the bar.[82]

Rather than suggesting that small lawyers and law firms should avoid generative AI tools, the New York case—when considered in light of our own results and prior research—can and should serve as a cautionary tale against uncritically using generative AI to practice law.

---

[78] See https://www.dlapiper.com/en-us/news/2023/03/dla-piper-to-utilize-cocounsel-the-groundbreaking-ai-legal-assistant-powered-by-openai-technology

[79] See https://casetext.com/blog/law-firm-dla-piper-announces-casetext-cocounsel/

[80] https://www.law.com/international-edition/2023/09/21/macfarlanes-joins-list-of-firms-adopting-harvey-ai/?slreturn=20230930102435#:~:text=Macfarlanes%20follows%20in%20the%20footsteps,or%20exploring%20the%20tool%20further.

[81] See https://www.forbes.com/sites/lanceeliot/2023/10/17/prestigious-symposium-on-ai-lawyering-reveals-keen-insights-including-the-ardent-debate-on-whether-to-use-generative-ai-in-law-school-education/?sh=33882bb312e2 (remarks of Rob Hill, Akin Gump,

[82] See,e.g, https://apnews.com/article/artificial-intelligence-chatgpt-courts-e15023d7e6fdf4f099aa122437dbb59b; https://www.reuters.com/legal/new-york-lawyers-sanctioned-using-fake-chatgpt-cases-legal-brief-2023-06-22/#:~:text=New%20York%20lawyers%20sanctioned%20for%20using%20fake%20ChatGPT%20cases%20in%20legal%20brief,-By%20Sara%20Merken&text=NEW%20YORK%2C%20June%2022%20(Reuters,an%20artificial%20intelligence%20chatbot%2C%20ChatGPT.

There are numerous well-known risks that come along with using generative AI as a tool for legal analysis, and the lawyers in that case ignored all of them. But small lawyers and law firms that interpret this incident to suggest the need to avoid generative AI reach precisely the wrong conclusion. Like any other tool, generative AI can be misused. The lesson to draw from this case, when considered in concert with the results of this study and prior evidence, is that lawyers and law firms that use generative AI tools must develop systems and procedures for doing so effectively. At the very least, these systems should include (i) confirming the veracity of any factual statements or characterizations of legal source materials made by AIs, (ii) experimenting with different prompting strategies when using general purpose AIs, including few-shot and grounded prompting, (iii) assessing legal issues and tasks independently of AI, and (iv) avoiding entering any confidential information into general purpose AIs that do not include trustworthy assurance of confidentiality.[83]

3. Legal Clients

The potential for generative AI to significantly improve the efficiency of legal work should be welcome news to many clients, particularly in the transactional domain. But rather than relying on market forces alone to decrease the cost of legal work product or increase the quality, we believe that our results suggest that clients should be proactive in asking their attorneys how they make use of generative AI and what impact that has on the quality and cost of the resulting legal services.

Despite the fiduciary nature of the attorney-client relationship, like all principal-agent relationships this relationship is characterized by various potential conflicts of interest.[84] Chief among them, of course, is the incentive of lawyers to spend more time performing legal work so as to increase the fees that they can charge.[85] Some lawyers may be inclined to accomplish this simply be resisting incorporating generative AI into their workflows, citing some of the risks of this technology described above. Others may explain to clients that their use of generative AI has allowed them to invest their scarce time into other ways of protecting the

[83] See Schwarcz & Choi, Practical Tips, supra.

[84] See Dennis M. O'Dea, The Lawyer-Client Relationship Reconsidered: Methods for Avoiding Conflicts of Interest, Malpractice Liability, and Disqualification, 48 Geo. Wash. L. Rev. 693, 730-32 (1980).

[85] See Lisa G. Lerman, A Double Standard for Lawyer Dishonesty: Billing Fraud Versus Misappropriation, 34 Hofstra L. Rev. 847, 848 (2006)

clients' interests. Of course, how convincing these answers are will depend on innumerable factors; but many clients who do not closely monitor how their lawyers' legal work product and billing practices are impacted by generative AI may end up paying more for less relative to their competitors.

An alternative approach for legal clients is to shift the balance of work that is outsourced to law firms rather than being produced in house.[86] The efficiencies associated with generative AI are virtually certain to shift the calculations associated with this make-buy decision. Most obviously, generative AI should allow clients to complete a larger percentage of routine legal work in house. Additionally, the uncertainty that generative AI introduces in how long legal work should take also counsels in favor of moving relatively routine work from external counsel to in house, as that shift should allow firms to better calibrate these expectations internally, where principal-agent problems are reduced.

These dynamics may well play out differently in adversarial settings, like high-stakes litigation. In litigation, both plaintiffs and defendants can use generative AI tools to increase the efficiency with which they produce relevant work product. As such, it is not clear that these efficiencies can or will result in an overall reduction in the optimal amount of time necessary to litigate a case, given the expectation that this technology may free up time for one's opponent to strengthen their case. Similar dynamics apply to fields like transactional contract negotiation, where AI might simply allow both sides to a deal to dig deeper and create ever-more-detailed contracts. In other words, competitive dynamics make it harder for clients to calibrate how access to generative AI should impact their legal bills, particularly with respect to domains high-stakes litigation or corporate mergers and acquisitions where outcomes matter much more than the size of the legal bills.

4. Judges

In the wake of the notorious New York case described above, a number of judges have prohibited lawyers that practice before them from using generative AI to assist with writing briefs.[87] In our view, our

[86] See John Armour & Mari Sako, *AI- Enabled Business Models In Legal Services: From Traditional Law Firms To Next-Generation Law Companies*, 7 Journal of Professions and Organization 27 (2020).

[87] Megan Cerullo, Texas judge bans filings solely created by AI after ChatGPT made up cases, Moneywatch, June 2, 2023, at https://www.cbsnews.com/news/texas-judge-bans-chatgpt-court-filing/; Judge Prohibits Out-Of-State Lawyer From Using ChatGPT,

results suggest that this approach is misguided; generative AI has the capacity to allow lawyers to better serve their clients by producing work product more efficiently, thus reducing barriers to justice. While generative AI can indeed be used irresponsibly to produce fabricated citations or source material, that risk can be safeguarded against through ordinary tools available to judges, such as their ability to impose Rule 11 sanctions. Resorting instead to an outright ban on the use of generative AI by lawyers because of one high-profile misuse of this technology undermines its ability to benefit clients and lawyers alike.

V. CONCLUSION

We conduct the first randomized controlled trial evaluating how LLMs can assist with legal analysis. We find small and variable improvements to the quality of work product but large and consistent improvements to speed. Moreover, we find that when AI provides a boost to quality at all, the boost to quality (but not speed) is inversely correlated with baseline performance, with a substantial improvement for the worst performers but essentially no improvement for the best. Finally, we find that participants accurately perceived how useful AI assistance was on each task and reported positive impressions from using AI at legal tasks. These findings suggest that AI could substantially transform the legal profession, streamlining tasks, improving lawyer satisfaction, and reducing inequality between lawyers.

https://www.law360.com/articles/1694884/judge-prohibits-out-of-state-lawyer-from-using-chatgpt

APPENDIX

A. Randomization Checks

To validate that Group A and Group B were correctly randomized, we compare whether the two groups match on observables. We collected individual-specific data for class year and 1L Fall GPA. We did not collect other demographic information out of concerns about anonymity. Table __ provides information about individual characteristics, including means and standard deviations, as well as the difference between the two groups. The differences have a *p*-value of 0.44 for class year and 0.92 for 1L Fall GPA, far from statistically significant.

Table 5: Group A and Group B Individual Characteristics

	Group A	Group B	Difference (95% CI)
Class Year	2024.38	2024.52	0.14
	(0.68)	(0.69)	(-0.48, 0.21)
1L Fall GPA	3.35	3.34	0.01
	(0.36)	(0.35)	(-0.17, 0.19)

In addition, we conducted Kolgorov-Smirnov tests to estimate the likelihood that the class years for Group A and Group B, and the 1L Fall GPAs for Group A and Group B, were drawn from the same distribution. The Kolgorov-Smirnov statistic for class year was 0.14 ($p = 0.95$) and for 1L Fall GPA was 0.10 ($p = 1.00$), again suggesting no difference between the two groups.

B. Graphs of Differences in Means

The following Figures show the distribution of differences in mean grade on each task, as well as the differences in the time taken for each task, between the group with and without access to GPT. The distributions were generated by calculating means on bootstrapped distributions, with 10,000 iterations.

Figure 18: Difference in Grade with Access to AI—Complaint Drafting

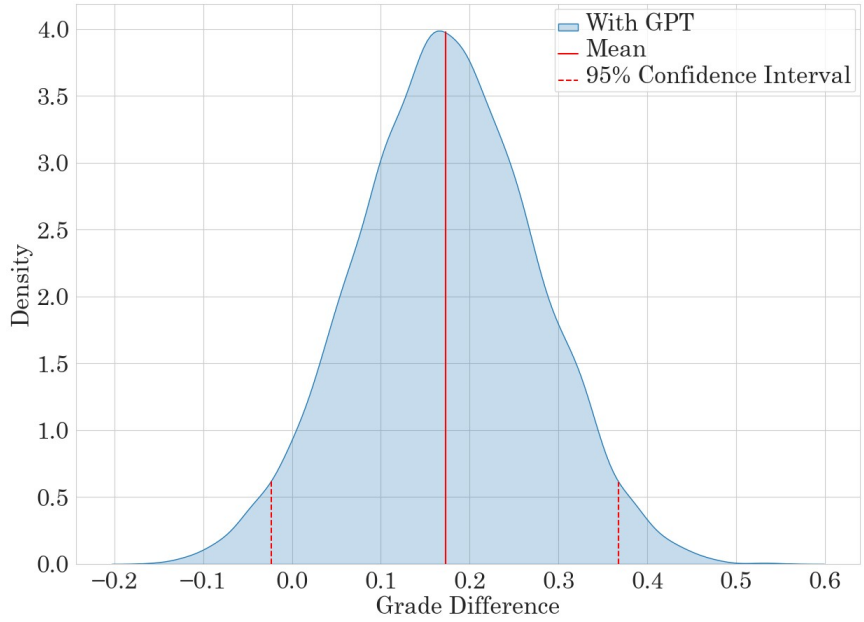


Figure 19: Difference in Grade with Access to AI—Contract Drafting

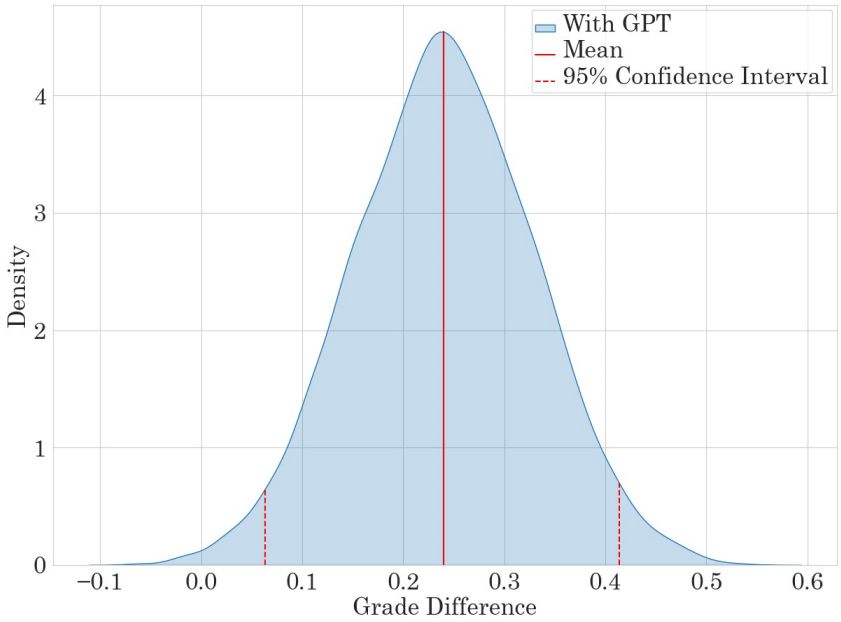


Figure 20: Difference in Grade with Access to AI—Employee Handbook

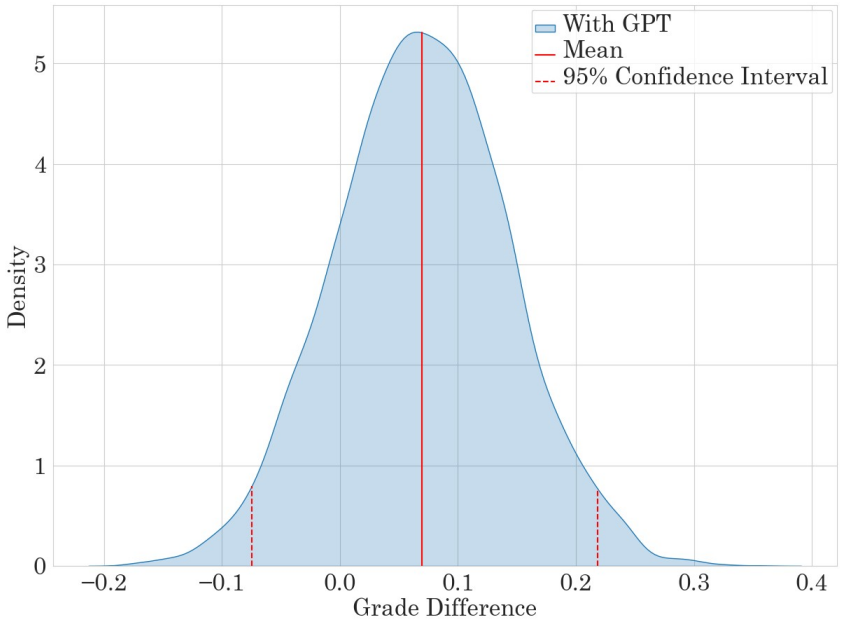


Figure 21: Difference in Grade with Access to AI—Client Memo

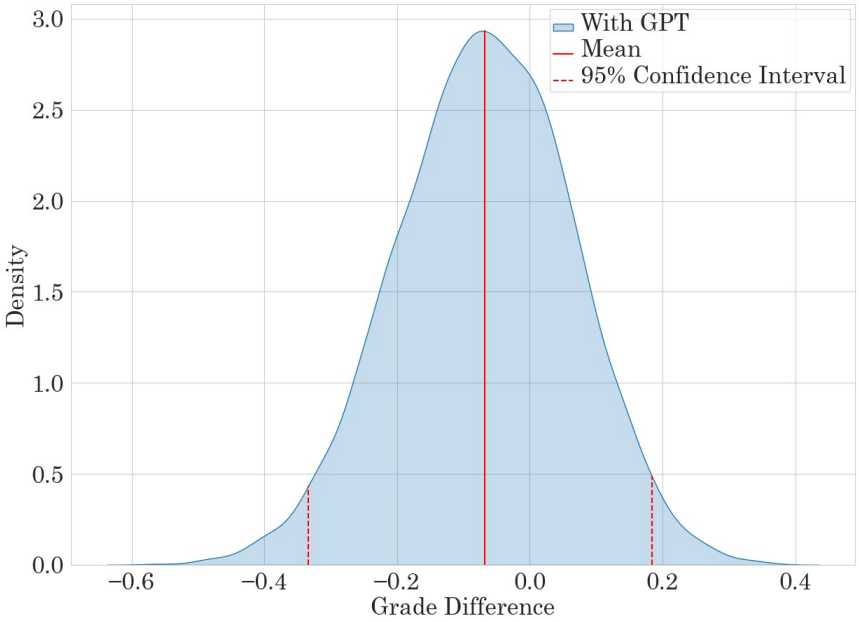


Figure 22: Difference in Time Taken with Access to AI—Complaint Drafting

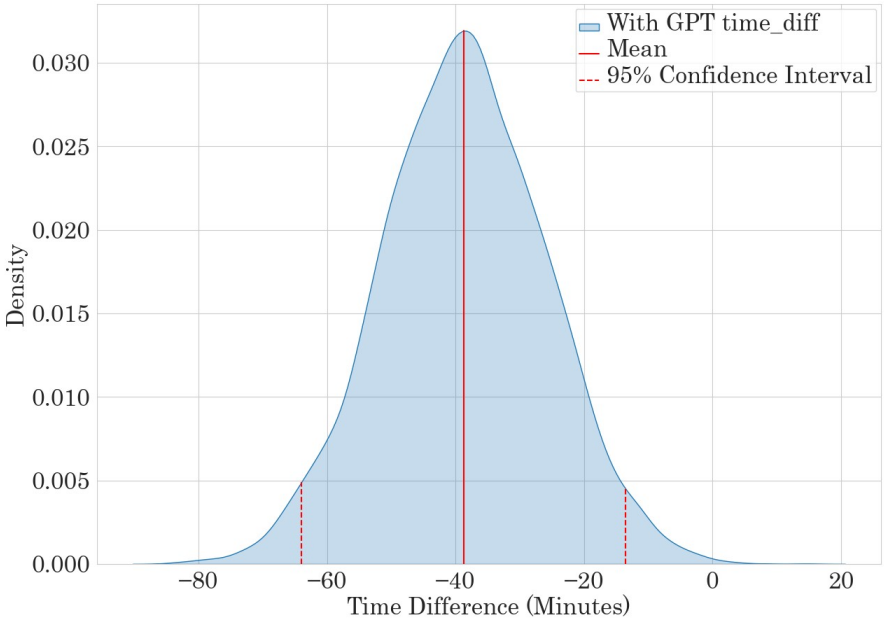


Figure 23: Difference in Time Taken with Access to AI—Complaint Drafting

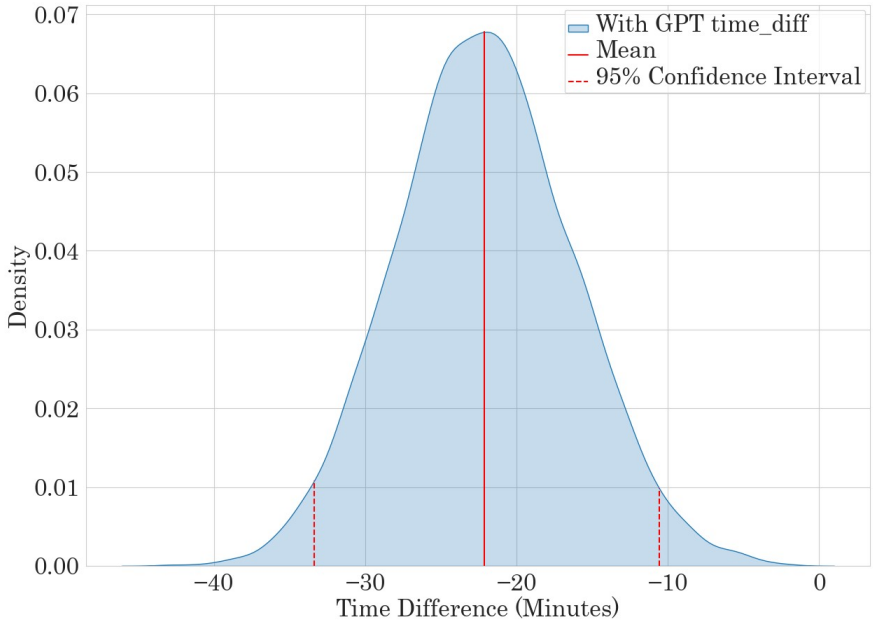


Figure 24: Difference in Time Taken with Access to AI—Employee Handbook

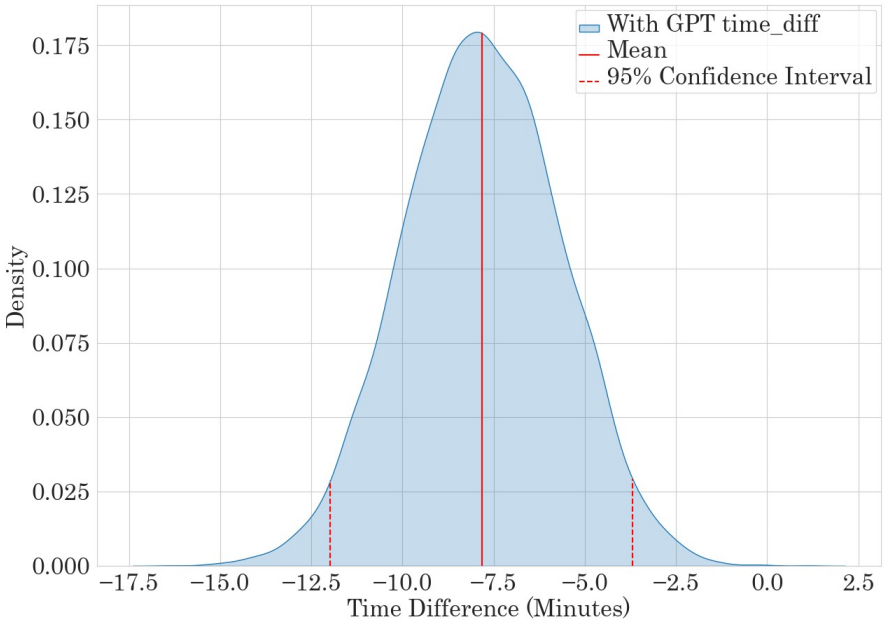
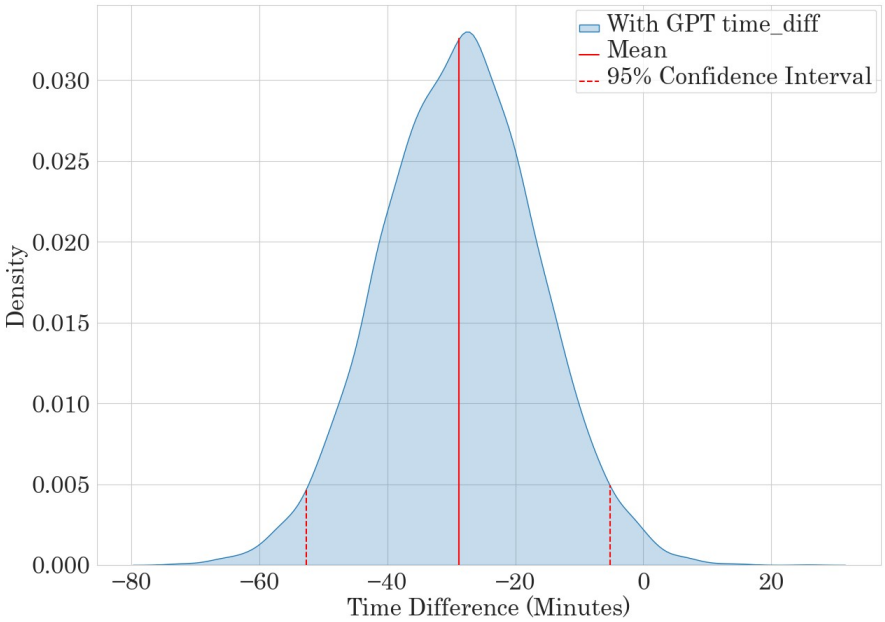


Figure 25: Difference in Time Taken with Access to AI—Client Memo



C. Training Materials

Prior to completing the four required tasks, participants completed an online training module that we developed and taught on how to use GPT-4 effectively in legal analysis.[88] This training involved watching three pre-recorded videos, totaling approximately two hours in length, and completing several short exercises requiring the use of GPT-4 to answer simple legal questions.[89] Training was split into three sub-areas. The first covered general principles on using AI effectively in legal research and writing.[90] Among other things, it provided participants with an overview of basic prompting techniques that prior research had shown to be effective in the legal setting, such as supplying the AI with relevant

[88] This training drew heavily on previous work by two of us. See AI Tools for Lawyers: A Practical Guide. Daniel Schwarcz & Jonathan H. Choi, *AI Tools for Lawyers: A Practical Guide*, 108 Minn. L. Rev. Headnotes (forthcoming 2023).

[89] Most people can access GPT-4 by creating a paid ChatGPT Plus account on the OpenAI website. However, it was not administratively possible to create such an account for each study participant without requiring participants to outlay cash on the subscriptions themselves. We instead created a central ChatGPT "clone" website using the GPT-4 API and gave students access to that website. This clone website had a nearly identical user interface and used the same system prompt as the real ChatGPT Plus.

[90] These general principles included the following key pieces of advice: (i) think about any legal problem first—develop your own basic instincts about key issues, principles, and parameters of work product you will need to produce; (ii) Start prompts by giving AI context that it should use to approach a question (i.e. "You are an experienced litigator"); (iii) Use AI to refine initial assessment of project by asking it to produce an outline, identify key issues, or produce first draft (in case of shorter assignments); (iv) Chunk up elements of outline, issues, application of rules into bite-sized bits, and ask AI to analyze each bit; adjust level of generality based on problem, quality of answers; (v) Provide AI with all the key details that a person would need to accomplish prior step; (vi) Iterate by providing additional details that you may have left out, asking AI to alter elements that do not look good, or asking AI to elaborate on elements that do look promising; (vi) Provide AI with relevant source materials, including cases, statutes, contract parameters, etc.; (vii) Do not rely on AI to conduct specific legal research or identify specific legal source material unless you confirm veracity of that material.

legal rules or source materials within prompts.[91] Second, the training covered basic techniques for using AI effectively in litigation-oriented settings, covering topics such as using AI to summarize and apply primary sources like caselaw and statutes.[92] The third and final portion of the training focused on using AI to draft transaction-oriented work product, such as contracts, highlighting AI's capacity to mimic conventional format, style, and structure of sample transactional

[91] See Jonathan H. Choi & Daniel Schwarcz, AI Assistance in Legal Analysis: An Empirical Study, Journal of Legal Education (forthcoming, 2024) https://ssrn.com/abstract=4539836. For a review of the computer science literature on these prompting strategies, see *ee, e.g.*, Dils, *How to Use ChatGPT: Advanced Prompt Engineering*, WGMI Media (Jan. 31, 2023), https://wgmimedia.com/how-to-use-chatgpt-advanced-prompt-engineering; *Awesome ChatGPT Prompts*, GitHub, https://github.com/f/awesome-chatgpt-prompts/#readme; Alan D. Thompson, *Microsoft Bing Chat (Sydney/GPT-4),* Life Architect (Feb. 22, 2023), https://lifearchitect.ai/bing-chat; Tyler Cowen & Alexander T. Tabarrok, How to Learn and Teach Economics with Large Language Models, Including GPT (Mar. 17, 2023) (unpublished manuscript) (on file with authors). *See also AI and Machine Learning Experts, Experienced Attorneys, Thousands of Hours of Prompt Engineering—and That's Just to Launch*, CaseText (May 12, 2023), https://casetext.com/blog/building-an-ai-legal-assistant-lawyers-can-trust. Jason Wei et al., *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models*, *in* Proceedings of the 36th International Conference on Neural Information Processing Systems 4356 (2022). Tom B. Brown et al., *Language Models Are Few-Shot Learners*, *in* Advances in Neural Information Processing Systems 33 (2020); Baolin Peng et al., Check Your Facts and Try Again: Improving Large Language Models with External Knowledge and Automated Feedback (Mar. 8, 2023) (unpublished manuscript) (on file with authors).

[92] This training suggested that participants (i) Independently review source material briefly, (ii) Ask GPT-4 to summarize specific cases and statutes by copying and pasting that material into GPT4 (and breaking it up into chunks if it is too long; (iii) Ask GPT-4 any relevant follow-up questions focusing in on elements of reasoning, issues, or facts that are most relevant., (iv) Ask GPT-4 to quote from the relevant source material in any of its explanations so you can verify it, and (v) Use GPT-4 to analogize or distinguish cases to specific fact pattern/scenario, highlighting key issues.

materials and to help identify alternative terms, unanticipated risks, and ambiguities in initial drafts.[93]

[93] More specifically, this portion of the training emphasized that AI can help (i) mimic the format/style/structure of any sample transactional material, (ii) incorporate specific deal terms or parameters into transactional documents if you provide those terms, (iii) identify potential risks to address, ambiguities in deal terms, (iv) help you issue spot potential additional terms to add to an agreement, and (v) help you further develop/specify terms, or identify alternative ways of drafting that can favor one particular side in the transaction.