Article

Lawyering in the Age of Artificial Intelligence

Jonathan H. Choi,† Amy B. Monahan,†† and Daniel Schwarcz†††

We conducted the first randomized controlled trial to study the effect of AI assistance on human legal analysis. We randomly assigned law school students to complete realistic legal tasks either with or without the assistance of GPT-4, tracking how long the students took on each task and blind-grading the results.

We found that access to GPT-4 only slightly and inconsistently improved the quality of participants' legal analysis but induced large and consistent increases in speed. AI assistance improved the quality of output unevenly—where it was useful at all, the lowest-skilled participants saw the largest improvements. On the other hand, AI assistance saved participants roughly the same amount of time regardless of their baseline speed. In follow-up surveys, participants reported increased satisfaction from

† Professor of Law, University of Southern California Gould School of Law.

†† Distinguished McKnight University Professor and Melvin C. Steen Professor, University of Minnesota Law School.

††† Fredrikson & Byron Professor of Law, University of Minnesota Law School. For their feedback, we thank David Abrams, Ben Alarie, Paul Connell, Kristin Hickman, David Hoffman, Michael Morse, Anthony Niblett, Jed Stiglitz, Eric Talley, and the participants at the Mitchell Hamline Faculty Workshop, the University of Pennsylvania Law and Economics Seminar, the Southern Methodist University Conference on ChatGPT, the University of Toronto Law School Faculty Workshop, the Harvard Law School Legal Profession Seminar, and the Duke Law School Faculty Workshop. This research was supported by the Kommerstad Faculty Imagination Fund at the University of Minnesota Law School.

using AI to complete legal tasks and correctly guessed the tasks for which GPT-4 was most helpful.

These results have important descriptive and normative implications for the future of lawyering. Descriptively, they suggest that AI assistance can significantly improve productivity and satisfaction, and that it can be selectively employed by lawyers in areas where AI is most useful. Because AI tools have an equalizing effect on performance, they may also promote equality in a famously unequal profession. Normatively, our findings suggest that law schools, lawyers, judges, and clients should thoughtfully embrace AI tools and plan for a future in which they will become widespread.

TABLE OF CONTENTS

INTRODUCTION

Rapid new improvements in the performance of artificial intelligence (AI) models have triggered excitement and trepidation about the future of lawyering.[1] Will AI replace human lawyers, or will it make them happier and more efficient? Should lawyers and judges embrace AI to perform legal tasks, or should they eschew it as unreliable and opaque? Should law schools incorporate AI into the curriculum, or is AI too speculative to be worth learning about?

Studies to date offer limited insight into these questions. Existing scholarship focuses on AI's ability to conduct legal analysis on its own, rather than its ability to assist humans.[2] Yet the

---

1. *See* John G. Roberts, Jr., *2023 Year-End Report on the Federal Judiciary*, SUP. CT. OF THE U.S. 5 (2023), https://www.supremecourt.gov/publicinfo/year-end/2023year-endreport.pdf [https://perma.cc/T6SC-FH9S] (urging "caution and humility" in the use of artificial intelligence); Roger Barton, *How Will Leveraging AI Change the Future of Legal Services?,* REUTERS (Aug. 23, 2023), https://www.reuters.com/legal/legalindustry/how-will-leveraging-ai-change-future-legal-services-2023-08-23 [https://perma.cc/6DYK-3END] (arguing that lawyers must adapt to, rather than compete with, AI advancements); Daniel Farrar, *To Future-Proof Their Firms, Attorneys Must Embrace AI*, FORBES (July 13, 2023), https://www.forbes.com/sites/forbesbusinesscouncil/2023/07/13/to-future-proof-their-firms-attorneys-must-embrace-ai/?sh=6282438b245b [https://perma.cc/SHK5-FHLN] (recommending lawyers to leverage and not fear AI); Steve Lohr, *A.I. Is Coming for Lawyers, Again*, N.Y. TIMES (Apr. 10, 2023), https://www.nytimes.com/2023/04/10/technology/ai-is-coming-for-lawyers-again.html [https://perma.cc/5D7C-HDGU] (emphasizing the simultaneous threat and opportunity of AI in the legal industry); John Villasenor, *How AI Will Revolutionize the Practice of Law,* BROOKINGS INST. (Mar. 20, 2023), https://www.brookings.edu/articles/how-ai-will-revolutionize-the-practice-of-law [https://perma.cc/C3MU-6E6L] (evaluating the future role of AI in the legal field).

2. *See, e.g.*, Jonathan H. Choi, Kristin E. Hickman, Amy B. Monahan & Daniel Schwarcz, *ChatGPT Goes to Law School*, 71 J. LEGAL EDUC. 387, 388–91 (2022) (finding that exam answers drafted by ChatGPT, with limited prompt engineering, achieved an average grade of a C+ in four real exams at the University of Minnesota Law School); Daniel Martin Katz, Michael James Bommarito, Shang Gao & Pablo Arredondo, *GPT-4 Passes the Bar Exam*, PHIL. TRANS. R. SOC. A, Apr. 15, 2024, at 1, 3–5 (finding that under zero-shot performance—where a model completes a task without training examples—GPT-4 passed the Uniform Bar Examination and outperformed the average human test-taker by seven percent on the Multistate Bar Examination portion); Matthew Dahl, Varun Magesh, Mirac Suzgun, & Daniel E. Ho, *Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models*, 16 J. LEGAL ANALYSIS 64, 66 (2024) (finding a "widespread occurrence of legal hallucinations" in analysis conducted by large language models, but failing to account for the

latter application is significantly more plausible for the foreseeable future given lawyers' ethical obligation to ensure that their work product is accurate and consistent with their clients' interests,[3] as well as the irreducibly normative nature of law.[4] A second limitation of prior research is that, for reasons of convenience, it has generally focused on how AI impacts performance on exams, like law school exams and the bar exam.[5] But exam results may not translate to lawyering in the real world.[6] Finally, many studies to date have suffered from methodological

straightforward methods to ameliorate hallucinations that we discuss in our training materials).

3. *See, e.g.*, The State Bar of Cal. Standing Comm. on Pro. Resp. and Conduct, *Practical Guidance for the Use of Generative Artificial Intelligence in the Practice of Law*, THE STATE BAR OF CAL. 3 (Nov. 17, 2023), https://www.calbar.ca.gov/Portals/0/documents/ethics/Generative-AI-Practical-Guidance.pdf [https://perma.cc/JG6S-J62V] ("A lawyer must critically review, validate, and correct both the input and the output of generative AI to ensure the content accurately reflects and supports the interests and priorities of the client in the matter at hand . . . ."); Jonathan Grabb, *Lawyers and AI: How Lawyers' Use of Artificial Intelligence Could Implicate the Rules of Professional Conduct*, THE FLA. BAR (Mar. 13, 2023), https://www.floridabar.org/the-florida-bar-news/lawyers-and-ai-how-lawyers-use-of-artificial-intelligence-could-implicate-the-rules-of-professional-conduct [https://perma.cc/79BH-JUGZ] ("While a chatbot may be able to draft a document in mere seconds, any lawyer who uses AI assistance is still responsible for generating work product that is legally and factually accurate, competent, and meritorious."); *see also* Nicole Yamane, *Artificial Intelligence in the Legal Field and the Indispensable Human Element Legal Ethics Demands*, 33 GEO. J. LEGAL ETHICS 877, 882 (2020) (noting ethical concerns due to the rising use of AI); W. Bradley Wendel, *The Promise and Limitations of Artificial Intelligence in the Practice of Law*, 72 OKLA. L. REV. 21, 24–26 (2019) (differentiating between routine legal work that may be automated by AI and high-risk work that still requires the human element).

4. *See, e.g.*, Frank Pasquale, *A Rule of Persons, Not Machines: The Limits of Legal Automation*, 87 GEO. WASH. L. REV. 1, 6 (2019) (advocating for understanding "technology as a tool to complement attorneys' skills, rather than substitute for them"); Rebecca Crootof, Margot E. Kaminski & W. Nicholson Price II, *Humans in the Loop*, 76 VAND. L. REV. 429, 486 (2023) (describing efforts by lawyers and law professors to "keep[] human lawyers involved in legal processes rather than relying fully on AI").

5. *See infra* Part I (denoting the limited nature of GPT-4 studies in the legal field).

6. *See, e.g.*, Marsha Griggs, *Building a Better Bar Exam*, 7 TEX. A&M L. REV. 1, 2 (2019) (discussing challenges to how well performance on the bar exam measures "readiness to enter the legal profession"); JOAN W. HOWARTH, SHAPING THE BAR: THE FUTURE OF ATTORNEY LICENSING 99–109 (2023) (arguing for significant reforms in the bar exam because it has historically failed to test the skills that new lawyers need to represent clients while unfairly harming traditionally marginalized groups).

limitations, like non-blind grading of results[7] or imperfectly matched treatment and control groups.[8]

To better understand how AI will affect the lawyers of the future and what should be done now, we conducted the first randomized controlled trial of the effect of large language model (LLM) assistance on human legal analysis.[9] To do so, we randomly assigned sixty students at the University of Minnesota Law School to complete four separate legal tasks (resulting in 240 total task completions), either with or without the assistance of the most advanced general-purpose generative AI tool then available, GPT-4.[10] We selected the four assigned tasks—drafting a complaint, a contract, a section of an employee handbook, and a client memo—because they typify the type of work performed by junior attorneys.[11] Prior to completing these tasks, study participants received several hours of training on how to use GPT-4 effectively, which we patterned on real attorney training materials.[12] After participants completed the four assigned

7. *See* Katz, Bommarito, Gao, & Arredondo, *supra* note 2, at 9 (acknowledging that the answers produced by GPT-4 in the study were not blindly graded, but attempting to address this issue by soliciting the views of peers who were provided with blind samples of the answers produced by GPT-4). *See generally* Eric Martínez, *Re-Evaluating GPT-4's Bar Exam Performance*, A.I. & L., Mar. 30, 2023, at 1 (critiquing OpenAI's claims that GPT-4 performed at the ninetieth percentile on the Uniform Bar Examination).

8. *See* Jonathan H. Choi & Daniel Schwarcz*, AI Assistance in Legal Analysis: An Empirical Study*, 73 J. LEGAL EDUC. (forthcoming 2024) (manuscript at 29–35) (on file with the Minnesota Law Review) [hereinafter *AI Assistance in Legal Analysis*] (reporting that the impact on exam scores of providing students with access to GPT-4 depended significantly on the student's starting skill level, while acknowledging various methodological limitations in the study's approach to measuring this effect).

9. The raw data is on file with the authors and may be requested.

10. *See* OpenAI, *GPT-4 Technical Report*, ARXIV 14 (Mar. 4, 2024), https://arxiv.org/abs/2303.08774 [https://perma.cc/L4X2-KZC6] (reporting that GPT-4's performance on various benchmarks exceeds the performance of prior generative AI models).

11. *See* Ann Sinsheimer & David J. Herring*, Lawyers at Work: A Study of the Reading, Writing, and Communication Practices of Legal Professionals*, 21 LEGAL WRITING: J. LEGAL WRITING INST. 63, 99–100 (2016) (reporting the results of a three-year ethnographic study of junior associates at law firms, which found that common documents that these lawyers drafted included formal summaries of their research findings, contracts, and complaints, among many other documents).

12. *See infra* Part II (describing the process for training study participants to use GPT-4 effectively to complete basic legal writing tasks).

tasks, we blind-graded the results and tracked how long they took on each task.[13]

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, and -0.07 on a 4.0 grading scale).[14] However, we found that AI assistance consistently induced large declines in the amount of time taken to complete tasks (-24.1%, -32.1%, -21.1%, and -11.8%).[15] The benefits of AI assistance were not evenly distributed; for 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).[16] 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.[17]

We also surveyed participants on their perceptions of how access to GPT-4 impacted their work on the assigned legal tasks.[18] We found that, with respect to the tasks on which GPT-4 was most useful, participants reported increased satisfaction from using it.[19] Although they completed the survey without knowing their results, participants also correctly understood GPT-4's strengths and weaknesses, reporting that they expected the improvements in speed to be greater than the improvements in quality and correctly identifying the tasks at which GPT-4 induced larger quality improvements.[20] This suggests that although the benefits from AI use may be inconsistent, participants generally correctly perceived the tasks at which it was most useful and could selectively use AI in situations where it provides the greatest benefits.[21]

---

13. *See infra* Part II.

14. *See infra* Table 1 (tabulating task performance with and without GPT-4).

15. *See infra* Table 2 (tabulating duration of tasks with and without GPT-4).

16. *See infra* Figures 9–12 (demonstrating higher GPA benefits for lower-skilled participants).

17. *See infra* Table 2 (tabulating time reductions from the use of GPT-4 on tasks).

18. *See infra* pp. 185–86 (presenting the survey questions).

19. *See* Figure 17 and accompanying text.

20. *Id.*

21. *Id.*

Taken together, these results point toward large potential productivity gains from AI assistance in the legal profession, especially by reducing the time taken to conduct legal analysis. They also suggest that AI could be a force to improve lawyer satisfaction.[22] Moreover, the results almost certainly serve as a lower-bound estimate on AI's capacity to improve the efficiency of legal services for three reasons. First, whereas our participants used the general purpose AI GPT-4 to assist them with assigned tasks, lawyers are increasingly gaining access to specialized generative AI tools that already offer better performance than GPT-4 on legal tasks.[23] Second, study participants only received a few hours of training on GPT-4 before completing assigned tasks,[24] whereas lawyers that use AI-based tools will continually refine their ability to skillfully use AI over the course of months or years. Finally, and perhaps most obviously, rapid AI innovation has continued since we conducted the experiment in the summer of 2023 and will likely do so for the foreseeable future.[25]

---

22. *See id.* (noting a boost to personal satisfaction when using GPT-4).

23. *See infra* notes 99–102 and accompanying text (discussing improved tools). For instance, LexisNexis just recently launched an AI legal assistant that is built into its general-purpose search engine. *See* Press Release, LexisNexis, LexisNexis Launches Lexis+ AI, a Generative AI Solution with Hallucination-Free Linked Legal Citations (Oct. 25, 2023), https://www.lexisnexis.com/community/pressroom/b/news/posts/lexisnexis-launches-lexis-ai-a-generative-ai-solution-with-hallucination-free-linked-legal-citations [https://perma.cc/ZQ69-RRVB]. Similarly, Thomson Reuters, the owner of Westlaw, recently acquired the firm Casetext in large part due to its generative AI capabilities. *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-tech-provider-casetext-650-mln-2023-06-27 [https://perma.cc/S43E-D3XY]. Westlaw is currently working to integrate at least some of these capabilities into its Westlaw Precision product. *See* Press Release, Thomson Reuters, Thomson Reuters Unveils Generative AI Strategy Designed to Transform the Future of Professionals (Nov. 1, 2023), https://www.thomsonreuters.com/en/press-releases/2023/november/thomson-reuters-unveils-generative-ai-strategy-designed-to-transform-the-future-of-professionals.html [https://perma.cc/VH4H-FD4S].

24. *See infra* Appendix A (describing the training undergone by participants).

25. *See, e.g.*, *What's the Future of Generative AI? An Early View in 15 Charts*, MCKINSEY & CO. (Aug. 25, 2023), https://www.mckinsey.com/featured-insights/mckinsey-explainers/whats-the-future-of-generative-ai-an-early-view-in-15-charts [https://perma.cc/T7LE-SQAD] (illustrating the pace of innovation in generative AI).

Especially when understood as a lower-bound estimate on AI's potential impact on lawyering, our results have important normative implications for actors across the legal services industry.[26] Lawyers and judges should affirmatively explore how to incorporate AI into their work, though AI's usefulness will vary by practice area, task, and the stakes of the underlying matters.[27] 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.[28] Law schools should reassess when and how law students are trained to use AI, and when and how access to that tool should be limited.[29]

We develop these results and implications in four parts. Part I briefly reviews both the evolution of legal technology and the state of the scholarly literature on how AI can impact lawyering and other knowledge-based tasks. Part II details our methodology, which employs a randomized controlled trial that allows us to make a strong causal inference about AI's impact on legal tasks. In Part III, we highlight and discuss our key results, which demonstrate that generative AI can significantly improve the speed at which legal tasks are completed without degrading the quality of the resulting work product. The implications of these results are then discussed in Part IV, which emphasizes that virtually all actors in the legal ecosystem—including judges, lawyers, clients, law schools, and law students—should devote significant attention to ethically and intelligently incorporating generative AI into their daily workflows and into their broader decision-making. Finally, a technical Appendix includes additional details about our methodology and results.

I. BACKGROUND

The first legal databases were introduced fifty years ago, at the beginning of what many consider the modern era of legal

26. *See infra* Part V.B.
27. *See infra* Parts V.B.1, V.B.3.
28. *See infra* Part V.B.2.
29. *See infra* Part V.B.4.

technology.[30] 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.[31] 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.[32]

Historically, these major legal tech innovations have improved lawyer efficiency rather than fundamentally altered the core skills needed to be an effective lawyer.[33] 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.[34] But the skill involved in analyzing and applying cases and statutes remains fundamentally the same. Similarly, e-discovery tools allow lawyers to automate the search function in discovery,[35] but cannot

30. William G. Harrington, *A Brief History of Computer-Assisted Legal Research*, 77 LAW LIBR. J. 543, 553 (1984) (recounting the development of LexisNexis and Westlaw); Olufunmilayo B. Arewa, *Open Access in a Closed Universe: Lexis, Westlaw, Law Schools, and the Legal Information Market*, 10 LEWIS & CLARK L. REV. 797, 816 (2006) (describing the creation of online legal databases); James A. Sprowl, *Computer-Assisted Legal Research: Westlaw and Lexis*, 62. A.B.A. J. 320, 321–23 (1976) (describing the similarities and differences of the Lexis and Westlaw databases).

31. *See, e.g.*, Roberts, *supra* note 1, at 2–5 (describing the legal profession's adoption of technologies ranging from personal computers to digitalization and technology-assisted review of discovery-related documents).

32. *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) (highlighting the demand for the "computerization of legal services" in America).

33. *Compare* Mark Fenwick, Wulf A. Kaal & Erik P.M. Vermeulen, *Legal Education in the Blockchain Revolution*, 20 VAND. J. ENT. & TECH. L. 351, 357 (2017) (describing the introduction of "Legal Tech" tools that have increased the efficiency of lawyers), *with* Symposium, *Legal Reasoning and Artificial Intelligence: How Computers "Think" Like Lawyers*, 8 U. CHI. L. SCH. ROUNDTABLE 1, 21 (2001) (noting that AI systems available at the time were not capable of the type of analogical reasoning that lawyers and judges engage in).

34. *See* Raymond H. Brescia, Walter McCarthy, Ashley McDonald, Kellan Potts & Cassandra Rivais, *Embracing Disruption: How Technological Change in the Delivery of Legal Services Can Improve Access to Justice*, 78 ALB. L. REV. 553, 567–68 (2014) (attributing an increase in legal research speed to websites like "LexisNexis, Westlaw, and Bloomberg").

35. *See* John O. McGinnis & Russell G. Pearce, *The Great Disruption: How Machine Intelligence Will Transform the Role of Lawyers in the Delivery of Legal*

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. For example, AI tools like predictive coding in e-discovery systems have become increasingly prominent in recent years.[36] These tools allow a lawyer to code a sample of discovery documents, which are then used by an algorithm to identify other relevant documents.[37] To a certain degree, tools such as these actually displace an attorney's work.[38]

With each new innovation, lawyers have typically fretted about the implications for the legal profession and lawyer jobs.[39]

Services, 82 FORDHAM L. REV. 3041, 3047–48 (2014) (recounting the use of "predictive coding" by lawyers to automate e-discovery in part).

36. *Id.*

37. *See id.* at 3047; *see also* 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, 936 (2013) (arguing that the prediction of legal outcomes is of primary interest to potential clients, which can be assisted with the aid of technology that leverages data about similar legal questions).

38. *See* Maura R. Grossman & Gordon V. Cormack, *Quantifying Success: Using Data Science to Measure the Accuracy of Technology-Assisted Review in Electronic Discovery* (finding that "technology-assisted review" systems in e-discovery provided "significantly superior precision" compared to manual review), *in* DATA DRIVEN LAW: DATA ANALYTICS AND THE NEW LEGAL SERVICES 127, 150–51 (Ed Walters ed., 2019). *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").

39. *See, e.g.*, RICHARD SUSSKIND & DANIEL SUSSKIND, THE FUTURE OF THE PROFESSIONS: HOW TECHNOLOGY WILL TRANSFORM THE WORK OF HUMAN EXPERTS 66–67 (2015) (expressing the belief of the authors and other legal commentators that the legal world is "on the brink of unprecedented upheaval"); Katz, *supra* note 37, at 909 ("Welcome to law's information revolution—revolution already in progress." (footnote omitted)); Dana Remus & Frank Levy, *Can Robots Be Lawyers? Computers, Lawyers, and the Practice of Law*, 30 GEO. J. LEGAL ETHICS 501, 501 (2017) ("We assess frequently advanced arguments that automation will soon replace much of the work currently performed by lawyers."); Tanina Rostain, *Robots Versus Lawyers: A User-Centered Approach*, 30 GEO. J. LEGAL ETHICS 559, 560 (2017) (identifying a recent trend of predictions that new legal technologies will push lawyers out of much of their current work and leave only narrow elaborate areas of law to humans); Sean Semmler & Zeeve Rose, *Artificial Intelligence: Application Today and Implications Tomorrow*, 16 DUKE L. & TECH. REV. 85, 86 (2017) ("As artificial intelligence looms over the practice of law, it is important to dispel the notion that artificially-

If technology allowed the same work to be done in less time,[40] or could replace lawyers altogether for certain tasks,[41] it was feared that there would be fewer jobs available for lawyers. In some cases, lawyers have responded to these fears by employing self-regulatory tools to limit the permissible use of technologies that could undermine demand for legal services.[42] 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.[43] Moreover, it is possible that task automation could also

---

intelligent machines will replace humans."); Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 87, 87 (2014) (arguing that it is "overly broad" to conclude that AI will have little impact until it displays "higher-order cognition"); David C. Vladeck, *Machines Without Principals: Liability Rules and Artificial Intelligence*, 89 WASH. L. REV. 117, 117–20 (2014) (describing positive and negative visions of intelligent machines); 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 [https://perma.cc/2UHS-BWRY] (explaining new e-discovery software and its potential impact on the legal market); JAMES E. MOLITERNO, THE AMERICAN LEGAL PROFESSION IN CRISIS: RESISTANCE AND RESPONSES TO CHANGE 208 (2013) (detailing several states' struggles regarding confidentiality at the inception of email).

40. *See* Fenwick, Kaal & Vermeulen, *supra* note 33 (describing technology advances that led to efficiency gains in legal practice).

41. *See, e.g.*, Christopher A. Suarez, *Disruptive Legal Technology, COVID-19, and Resilience in the Profession*, 72 S.C. L. REV. 393, 404 (2020) (touting new e-discovery technology that "can eliminate the need for lawyers to review each and every document").

42. *See* Gillian K. Hadfield, *Legal Barriers to Innovation: The Growing Economic Cost of Professional Control Over Corporate Legal Markets*, 60 STAN. L. REV. 1689, 1724–25 (2008) (describing regulation of "unauthorized practice[s] of law" by state bar associations).

43. *See, e.g.*, SUSSKIND & SUSSKIND, *supra* note 39; McGinnis & Pearce, *supra* note 35, at 3055; Brescia, McCarthy, McDonald, Potts & Rivais, *supra* note 34, at 553 ("[T]he provision of legal services is becoming commodified: carried out by lawyers and nonlawyers alike in a way that is far less expensive than the traditional, 'bespoke' model of lawyering."). *See generally* Elinor R. Jordan, *Point, Click, Green Card: Can Technology Close the Gap in Immigrant Access to Justice?*, 31 GEO. IMMIGR. L.J. 287 (2017) (exploring the difficulties of obtaining a green card and ways that new legal technology has or can assist in the process); Kathleen Elliott Vinson & Samantha A. Moppett, *Digital Pro Bono: Leveraging Technology to Provide Access to Justice*, 92 ST. JOHN'S L. REV. 551 (2018) (identifying the need for more pro bono work and how that need can be facilitated by legal organizations and law schools employing new technologies); J.J. Prescott, *Improving Access to Justice in State Courts with Platform Technology*, 70 VAND. L. REV. 1993, 1993–94, 2026–39 (2017) (presenting evidence showing accessible state court websites reduce dispute resolution and fee payment times).

increase the demand for lawyers, either because the lower cost of legal services increases the overall quantity of legal services provided (induced demand) or because automation creates new tasks for which human labor is an important complement (what Daron Acemoglu and Pascual Restrepo have called a "reinstatement effect").[44]

Similar dynamics exist in recent discussions of how increasingly capable LLMs like GPT-4 will impact the legal profession.[45] At the same time, LLMs like GPT-4 seem 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,[46] radically changing market demand for human lawyers.[47]

Yet, despite these sizeable questions and concerns, relatively little is known empirically about AI's capacity to displace

---

44. Daron Acemoglu & Pascual Restrepo, *Automation and New Tasks: How Technology Displaces and Reinstates Labor*, 33 J. ECON. PERSPECTIVES 3, 3–4 (2019) (describing the countervailing actions of the "displacement effect," whereby labor is replaced by automation, and the reinstatement effect); *see also* Daron Acemoglu, David Autor, Johnathon Hazell, & Pascual Restrepo, *Artificial Intelligence and Jobs: Evidence from Online Vacancies*, 40 J. LABOR ECON. 293, 293 (2022) (analyzing the effect of AI on jobs but finding it too small for definitive conclusions); Daron Acemoglu, *The Simple Macroeconomics of AI* 23–42 (Nat'l Bureau of Econ. Rsch., Working Paper No. 32487, 2024) (analyzing the macroeconomic effects of recent AI developments).

45. *See, e.g.*, Erin Mulvaney & Lauren Weber, *End of the Billable Hour? Law Firms Get on Board with Artificial Intelligence*, WALL ST. J. (May 11, 2023), https://www.wsj.com/articles/end-of-the-billable-hour-law-firms-get-on-board-with-artificial-intelligence-17ebd3f8 [https://perma.cc/X6VG-UFK4] (discussing how AI tools can "handle the drudgery" of repetitive work and "simplify complex work").

46. 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"); Benjamin Alarie, Anthony Niblett & Albert H. Yoon, *How Artificial Intelligence Will Affect the Practice of Law*, 68 U. TORONTO L.J. 106, 114 (2018) (speculating that AI will substantially transform the work of lawyers in the future).

47. The impact of generative AI on the labor market is certainly not limited to the legal profession. *See, e.g.*, Tyna Eloundou, Sam Manning, Pamela Mishkin & Daniel Rock, GPTs Are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models (Aug. 22, 2023) (unpublished manuscript) (on file with the Minnesota Law Review) (evaluating the potential labor market effects of LLMs like GPT-4 on a variety of industries).

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,[48] bar examinations,[49] and in answering discrete legal questions.[50] Other non-empirical research considers the ethical implications of using such technology in the practice of law,[51] how artificial intelligence may change the skills needed to be a successful lawyer,[52] and how law firms may begin to compete on the basis of technological expertise.[53]

Studies examining GPT's proficiency on legal exams have found that its performance varies widely depending on the type of exam and prompting methodology used. One study found that GPT-4 alone performed in the ninetieth percentile on the

48. Choi, Hickman, Monahan & Schwarcz, *supra* note 2 (testing the performance of GPT-3.5 alone on law school exams); *see also* Andrew Blair-Stanek, Anne-Marie Carstens, Daniel Goldberg, Mark A. Graber, David Gray, & Maxwell L. Stearns, GPT-4's Law School Grades: Con Law C, Crim C-, Law & Econ C, Partnership Tax B, Property B-, Tax B, 2 (May 24, 2023) (unpublished manuscript) (on file with the Minnesota Law Review) [hereinafter GPT-4's Law School Grades] (reporting mixed results with law school exams).

49. *See* Katz, Bommarito, Gao & Arredondo, *supra* note 2, at 1, 7 (examining GPT-4's performance on different bar exam sections).

50. *See* John J. Nay et al., *Large Language Models as Tax Attorneys: A Case Study in Legal Capabilities Emergence*, PHIL. TRANSACTIONS ROYAL SOC'Y A, Apr. 15, 2024, at 1, 5–9 (studying large language models' ability to apply tax law); 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) [hereinafter *OpenAI Cribbed Our Tax Example*] (examining data from GPT-4 tax study).

51. *See, e.g.,* Katherine Medianik, Note, *Artificially Intelligent Lawyers: Updating the Model Rules of Professional Conduct in Accordance with the New Technological Era*, 39 CARDOZO L. REV. 1497, 1501 (2018) (examining the Model Rules of Professional Conduct in the context of AI). *See generally* Brian L. Frye, *Should Using an AI Text Generator to Produce Academic Writing Be Plagiarism?*, 33 FORDHAM INTELL. PROP., MEDIA & ENT. L.J. 946 (2023) (asking ChatGPT itself about plagiarism and the ethical considerations of its use).

52. *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, 1154 (2019) (describing the skills a lawyer will need in the context of AI); Suarez, *supra* note 41, at 396 (explaining that lawyers will need "increased resilience" in an "uncertain technological future").

53. *See* Bruce A. Green & Carole Silver, *Technocapital@BigLaw.com*, 18 NW. J. TECH. & INTELL. PROP. 265, 282–308 (2021) (studying technology and law firm competition).

Uniform Bar Examination[54] (although scholars have subsequently raised methodological doubts about this claim).[55] 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 appl[y] rules to facts as compared [to] real law students."[56]

---

54. Katz, Bommarito, Gao & Arredondo, *supra* note 2, at 12 n.3. 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 1. 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 correct[ing] the language" in the prompt so that it formed a complete sentence. *Id.* at 9.

55. *See, e.g.*, Martínez, *supra* note 7, at 16–17 (discussing potential methodological issues with the initial finding that GPT-4 surpassed the bar exam score of ninety percent 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. *Id.*

56. Choi, Hickman, Monahan & Schwarcz, *supra* note 2, at 393; *see also* GPT-4's Law School Grades, *supra* note 48, at 7–8 (finding that GPT-4 "performed well below average," despite being "decent[] at multiple choice" and sometimes spotting issues missed by students). There are concerns about how to format exams to address potential cheating. *See generally* Margaret Ryznar, *Exams in the Time of ChatGPT*, 80 WASH. & LEE L. REV. ONLINE 305 (2023).

In other disciplines, GPT has been found to be a proficient and sometimes superior test taker as compared to humans. *See* Harsha Nori, Nicholas King, Scott Mayer McKinney, Dean Carignan & Eric Horvitz, Capabilities of GPT-4 on Medical Challenge Problems 1 (Apr. 12, 2023) (unpublished manuscript) (on file with the Minnesota Law Review) (finding that GPT-4, without any specialized prompting, passes a range of medical exams and outperforms both ChatGPT and LLM models specifically fine-tuned on medical knowledge); John C. Lin, David N. Younessi, Sai S. Kurapati, Oliver Y. Tang & Ingrid U. Scott, *Comparison of GPT-3.5, GPT-4, and Human User Performance on a Practice Ophthalmology Written Examination*, 37 EYE 3694, 3695 (2023) ("GPT-4 but not GPT-3.5 achieved the passing threshold for a practice ophthalmology written examination."); Rohaid Ali et al., *Performance of ChatGPT and GPT-4 on Neurosurgery Written Board Examinations*, 93 NEUROSURGERY 1353, 1357–59 (2023) (finding that both GPT-4 and GPT-3.5 pass neurosurgery practice board exams at rates comparable to neurosurgery residents); Hanmeng Liu, Ruoxi Ning, Zhiyang Teng, Jian Liu, Qiji Shou & Yue Zhang, Evaluating the Logical Reasoning Ability of ChatGPT and GPT-4, at 5–6 (May 5, 2023) (unpublished manuscript) (on file with the Minnesota Law Review) (describing the performance of ChatGPT and GPT-4 on multi-choice reading comprehension and natural language inference datasets); Vinay Pursnani, Yusuf Sermet, Musa Kurt

Perhaps most interestingly, a later study examining GPT-4 *assistance* on law school exams, 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.[57]

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 to 70%–90% accuracy with significant human assistance (particularly prompting with hand-selected correct sources).[58] Many scholars have anecdotally tested GPT's capabilities, including a series of YouTube videos that illustrate GPT-4's capabilities in various legal contexts.[59] These 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.[60]

In areas other than law, we see the same general focus on exam performance rather than studies of realistic tasks.[61] And as with law, the exam results are mixed. Whereas exams

---

& Ibrahim Demir, Performance of ChatGPT on the US Fundamentals of Engineering Exam: Comprehensive Assessment of Proficiency and Potential Implications for Professional Environmental Engineering Practice 1 (Apr. 20, 2023) (unpublished manuscript) (on file with the Minnesota Law Review) (detailing the "significant improvement" of ChatGPT and GPT-4 in the context of the Fundamentals of Engineering Environmental Exam).

57. *AI Assistance in Legal Analysis*, *supra* note 8, at 12–13, 17.

58. Nay et al., *supra* note 50, at 8 (showing results with few shot and chain-of-thought prompting, in addition to gold truth retrieval); *cf. OpenAI Cribbed Our Tax Example*, *supra* note 50, at 1105 (describing GPT-4 using a simplified version of the Internal Revenue Code as a reference text and getting tax liabilities exactly correct roughly one-third of the time).

59. *See generally Harry Surden*, YOUTUBE (last visited Jan. 8, 2024), https://www.youtube.com/@harrysurden3116 (testing GPT-4's performance on various legal problems).

60. Harry Surden, *GPT-4 and Law: ChatGPT Analyzes Copyright Law*, YOUTUBE (Mar. 22, 2023), https://www.youtube.com/watch?v=nqZcrhR8yPU.

61. *See, e.g.*, Wayne Geerling, G. Dirk Mateer, Jadrian Wooten & Nikhil Damodaran, *ChatGPT Has Aced the Test of Understanding in College Economics: Now What?*, 68 AM. ECONOMIST 233, 237–39 (2023) (studying ChatGPT's performance on an economics exam).

generated by ChatGPT were highly rated in economics,[62] they achieved more middling results in computer programming ("outstanding to satisfactory")[63] and medical education,[64] and "unsatisfactory" results in fields like mathematics and psychology.[65] Common problems with ChatGPT-drafted exams included inaccurate, unreliable, and outdated information.[66] 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.[67] 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.[68]

---

62. *Id.* at 233 (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).

63. Chung Kwan Lo, *What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature*, 13 EDUC. SCIS., Apr. 2023, at 1, 5 (reviewing literature on ChatGPT's performance across subject areas, including computer programming).

64. Tiffany H. Kung et al., *Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models*, PLOS DIGITAL HEALTH, Feb. 2023, at 1 (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, 1238 (2023) (examining GPT-4's performance on various medical applications and finding that, although GPT-4 excels in some areas, it has important limitations).

65. Lo, *supra* note 63, at 6 (reviewing literature on ChatGPT's performance across subject areas and finding that its performance on mathematics and psychology problems is subpar); *see also* Lakshmi Varanasi, *ChatGPT Can Ace the Bar, but It Only Has a Decent Chance of Passing the CFA Exams. Here's a List of Difficult Exams the ChatGPT and GPT-4 Have Passed*, BUS. INSIDER (Nov. 5, 2023), https://www.businessinsider.com/list-here-are-the-exams-chatgpt-has-passed-so-far-2023-1 [https://perma.cc/SW5W-PDU2] (compiling study results and noting that GPT-4 "still struggles with high school math exams").

66. *See generally* Lo, *supra* note 63 (discussing the challenges of ChatGPT in education).

67. *See, e.g.*, Nori, King, McKinney, Carignan & Horvitz, *supra* note 56, at 1 (testing GPT acting alone).

68. *See, e.g.*, Fabrizio Dell'Acqua et al., *Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality* 2 (Harvard Bus. Sch., Working Paper 24-013, 2023) (on file with the Minnesota Law Review) [hereinafter *Navigating the Jagged Technological Frontier*] (studying the productivity of consultants working with GPT-4).

Outside of the exam setting, a small number of studies have evaluated how AI can improve human performance at non-legal professional writing tasks.[69] 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 to the least-skilled workers.[70] 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.[71]

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

69. 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, e.g.*, Sida Peng, Erini Kalliamvakou, Peter Cihon & Mert Demirer, The Impact of AI on Developer Productivity: Evidence from GitHub Copilot 2 (Feb. 13, 2023) (unpublished manuscript) (on file with the Minnesota Law Review) (examining "the productivity effects of AI on software development"). None of these studies evaluate how more-sophisticated prompting techniques can impact results.

70. Shakked Noy & Whitney Zhang, *Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence*, 381 SCI. 187, 190 (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. *Id.* at 187. 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. *Id.* After completing the first writing assignment, half of the participants were given access to ChatGPT for the second writing assignment. *Id.* 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. *Id.* at 188. 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. *Id.* at 188–89. 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. *Id*.

71. *See* Fabrizio Dell'Acqua, Falling Asleep at the Wheel: Human/AI Collaboration in a Field Experiment on HR Recruiters 1 (Dec. 2, 2021) (unpublished manuscript) (on file with the Minnesota Law Review) [hereinafter Falling Asleep at the Wheel] ("As AI quality increases, humans have fewer incentives to exert effort and remain attentive, allowing the AI to substitute, rather than augment their performance.").

management consulting.[72] 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.[73] 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.[74] Where consultants were working on problem-solving tasks that required the synthesis of quantitative data and qualitative information from interviews, AI provided much less of a boost.[75] 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.[76] 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.[77] However, in contrast to Choi & Schwarcz, the study found performance improvements even among those in the top half of skill rankings.[78] 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.[79] Researchers also found that participants who blindly adopted AI outputs

72. *Navigating the Jagged Technological Frontier*, *supra* note 68, at 2 (studying AI performance with Boston Consulting Group).

73. *Id.* at 8–15.

74. *Id.* at 7, 10; *see also* Karan Girotra, Lennart Meincke, Christian Terwiesch & Karl T. Ulrich, *Ideas are Dimes a Dozen: Large Language Models for Idea Generation in Innovation* 1 (Mack Inst. For Tech. Innovation, Working Paper, 2023) (on file with the Minnesota Law Review) (finding that GPT-4 can generate ideas faster and cheaper than college students at an elite university).

75. *Navigating the Jagged Technological Frontier*, *supra* note 68, at 13–15 (requiring GPT to use both quantitative and qualitative data resulted in "a significant negative impact" on GPT performance).

76. *Id.* at 10, 15.

77. *Id.* at 11 (finding a forty-three percent increase in performance among those ranked in the bottom half of skill level).

78. *Id.* at 11 (finding a seventeen percent increase in performance among those ranked in the top half).

79. *Id.* at 12; *see also* Leonard Boussioux, Jacqueline N. Lane, Miaomiao Zhang, Vladimir Jacimovic & Karim R. Lakhani, The Crowdless Future? How Generative AI Is Shaping the Future of Human Crowdsourcing 1 (Aug. 8, 2023) (unpublished manuscript) (on file with the Minnesota Law Review) (similarly finding that GPT-4 may decrease some forms of creativity and novelty compared to purely human outputs).

suffered a decrease in performance compared to those not using AI assistance at all.[80]

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.[81] GPT-4 and other LLMs sometimes hallucinate sources and sometimes fail to interpret sources accurately.[82] 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.[83]

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 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.[84] Well over 100 students expressed interest in participating in the study.[85] We initially enrolled the first sixty such

80. *Navigating the Jagged Technological Frontier*, *supra* note 68, at 17 ("Professionals who had a negative performance when using AI tended to blindly adopt its output and interrogate it less." (citation omitted)).

81. *See supra* notes 48–80 and accompanying text (detailing AI's capabilities and shortcomings).

82. *See* Dahl, Magesh, Suzgun & Ho, *supra* note 2, at 8–11 (describing hallucinations in the legal context); Nay et al., *supra* note 50, at 3, 8 (demonstrating inaccuracies when LLMs interpret tax law based on the CFR and the U.S. Code).

83. *See, e.g.*, *AI Assistance in Legal Analysis*, *supra* note 8, at 12–13, 17 (studying AI use on law exams).

84. The University of Minnesota Law School is one of the top law schools in the country, currently ranked sixteenth in the U.S. News ranking of law schools. *2024 Best Law Schools*, U.S. NEWS (2024), https://www.usnews.com/best-graduate-schools/top-law-schools/law-rankings [https://perma.cc/TG6H-9HER].

85. 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 fifteen hours total during June 2023. The email also noted that the work could be completed

volunteers and placed the remaining volunteers on a waitlist.[86] Over the duration of the study, twenty-two 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 fifty-nine students completed the experiment.

During the enrollment process, we gathered basic information about study participants, including their first-semester first-year law school GPA and their anticipated graduation year.[87] 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.

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.[88] 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

---

remotely and on participants' own time-schedules and that participants who completed the study would receive $300 in compensation for their time.

86. This experimental design was approved by the University of Minnesota's Institutional Review Board (IRB). Participants agreed to participate after reviewing and agreeing to an IRB-approved consent form.

87. We also collected contact information, including email and mailing addresses, and screened for prior enrollment in two classes that disqualified interested individuals from enrolling in the study because assigned study tasks overlapped with projects in those courses.

88. This training drew heavily on previous work by two of us. *See generally* Daniel Schwarcz & Jonathan H. Choi, *AI Tools for Lawyers: A Practical Guide*, 108 MINN. L. REV. HEADNOTES 1 (2023) [hereinafter *AI Tools for Lawyers*]. These materials have served as the basis for numerous practical training sessions that we have conducted for real lawyers in a variety of settings, including Continuing Legal Education presentations, presentations for in-house legal teams, and presentations for lawyers working at large law firms.

assignments on their own before using GPT-4 to generate answers. Additionally, the training required participants to practice these skills by using GPT-4 to answer sample problems. Section A 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 material 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 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. Section B of the Appendix contains additional information about these assignments.

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 discourage 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.[89] 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

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.

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.[90]

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 quality benefits, with little quality benefit to participants who were capable of producing high-quality work on their own. In contrast, the improvement in 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 without knowledge of their grades on the various tasks.

Table 1 below shows statistics for the grades received and time taken for each task.[91] It shows that the differences are relatively small in magnitude. Access to GPT-4 had the largest positive effect for contract drafting, where the difference in grade it

90. *See* Jonathan H. Choi, Amy B. Monahan & Daniel Schwarcz, *The Use of Artificial Intelligence to Assist with Basic Lawyering Tasks*, OPEN SCI. FRAMEWORK (Aug. 23, 2023), https://osf.io/n5saz [https://perma.cc/NJ29-G2V8]. *See generally* Jason M. Chin & Kathryn Zeiler, *Replicability in Empirical Legal Research*, 17 ANN. REV. L. SOC. & SCI. 239, 243 (2021) (discussing the benefits of pre-registering a data collection and analysis plan in the context of empirical legal research).

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

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 employee handbook task, respondents saw, on average, a near zero effect on performance from using GPT-4.

Table 1: Average Performance at Tasks with and Without GPT-4 (Grade on 4.0 Scale)

Task	No GPT-4 (Std. Dev.)	With GPT-4 (Std. Dev.)	Diff. (95% CI)	*p*-value
Complaint Drafting	3.14	3.31	0.17	0.0862
	(0.59)	(0.50)	(-0.03, 0.37)	
Contract Drafting	3.00	3.24	0.24	0.0060
	(0.56)	(0.40)	(0.07, 0.41)	
Employee Handbook	3.20	3.26	0.07	0.3532
	(0.41)	(0.39)	(-0.07, 0.21)	
Client Memo	2.92	2.85	-0.07	0.5980
	(0.69)	(0.76)	(-0.34, 0.18)	

Figure 1 through Figure 4 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).[92] Figure 18 through Figure 21 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 ninety-five percent level.

92. All figures in this Article were generated using the SciPy package in Python. Density plots were generated with Gaussian Kernel Density Estimation (used for irregular data distributions) through the gaussian_kde package in SciPy, which applies Scott's Rule to determine bandwidth. *See generally* Adriano Z. Zambom & Ronaldo Dias, *A Review of Kernel Density Estimation with Applications to Econometrics*, 5 INT'L ECONOMETRIC REV. 20, 29–33 (2013) (discussing methods of determining the smoothing parameter for kernel density estimation).

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

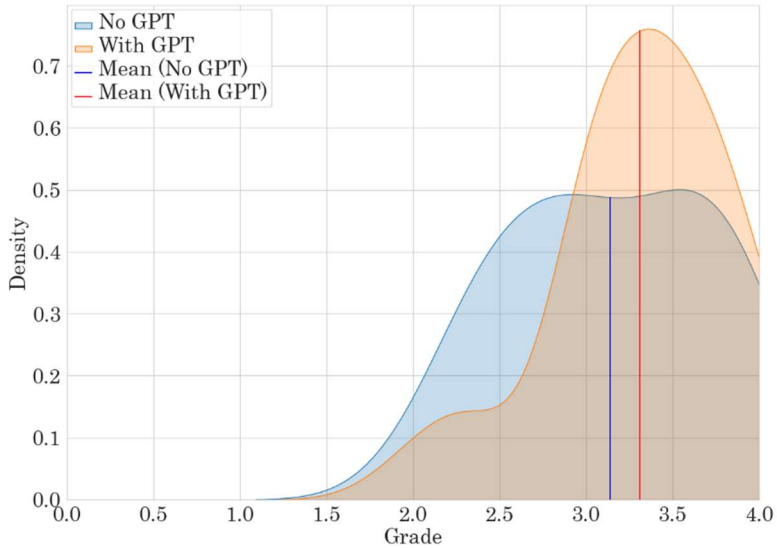


Figure 2: Quality Distributions with and Without AI—Contract Drafting

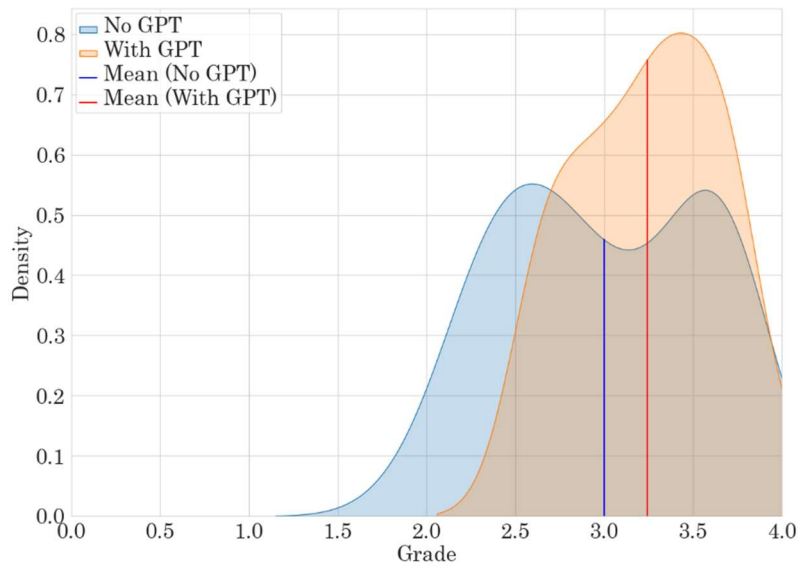


Figure 3: Quality Distributions with and Without AI—Employee Handbook

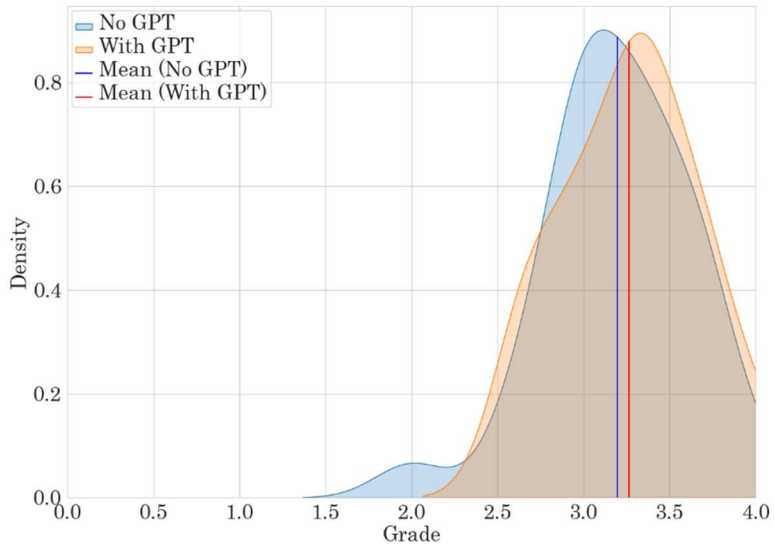


Figure 4: Quality Distributions with and Without AI—Client Memo

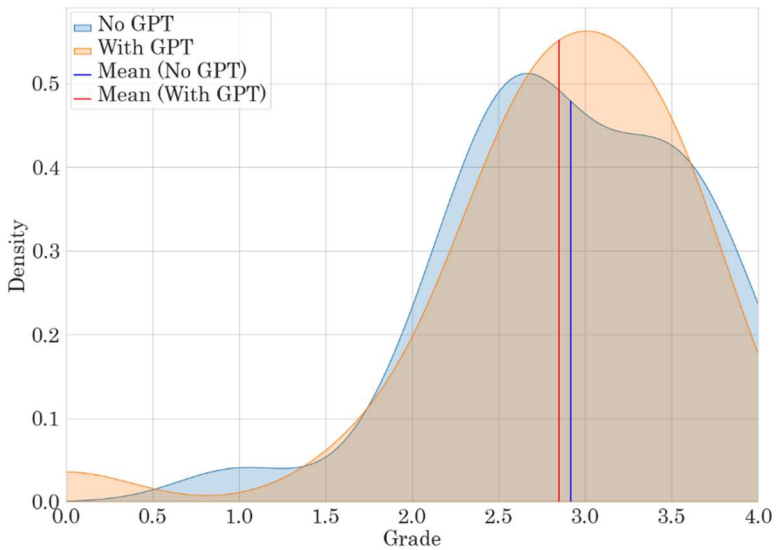


Table 2 below depicts the effect of access to GPT on the amount of time taken on each task. These results are more decisive, showing large and consistent decreases in the amount of time taken on each task. Interestingly, the largest gain in speed (in percentage terms) occurs in the task for which GPT-4 was the most useful in terms of grade improvement (contract drafting), and the smallest gain in speed occurs in the task for which GPT-4 was the least useful (client memo).

Table 2: Average Time Taken on Tasks with and Without GPT-4 (Minutes)

Task	No GPT-4 (Std. Dev.)	With GPT-4 (Std. Dev.)	Difference (95% CI)	% Diff.	*p*-value
Complaint Drafting	160.69	122.00	-38.77	24.1	0.0018
	(72.38)	(66.80)	(-64.00, -13.36)		
Contract Drafting	69.72	47.59	-22.40	32.1	0.0000
	(32.00)	(31.09)	(-33.71, -10.91)		
Employee Handbook	37.24	29.41	-7.84	21.1	0.0000
	(9.55)	(13.42)	(-12.03, -3.74)		
Client Memo	244.41	215.69	-28.75	11.8	0.0152
	(58.03)	(72.96)	(-52.59, -5.05)		

Figure 5 through Figure 8 below show the distributions of the amount of time that participants took on each task. Figure 22 through Figure 25 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 ninety-five percent level.

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

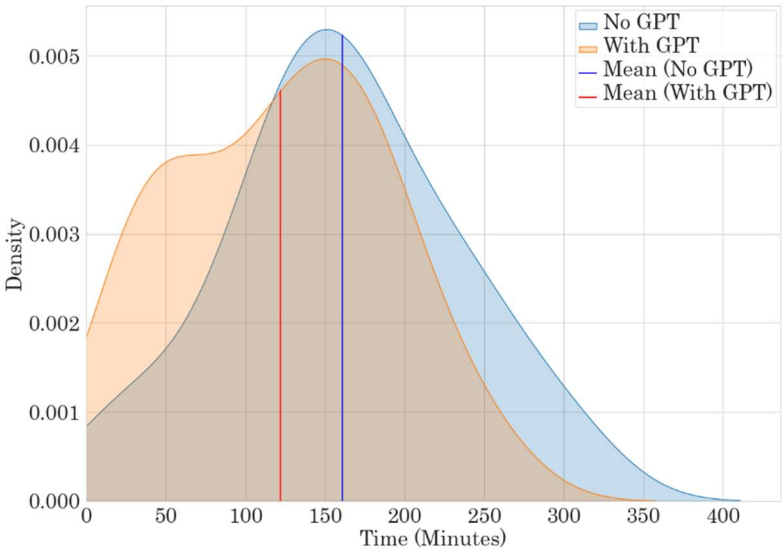


Figure 6: Time Distributions with and Without AI—Contract Drafting

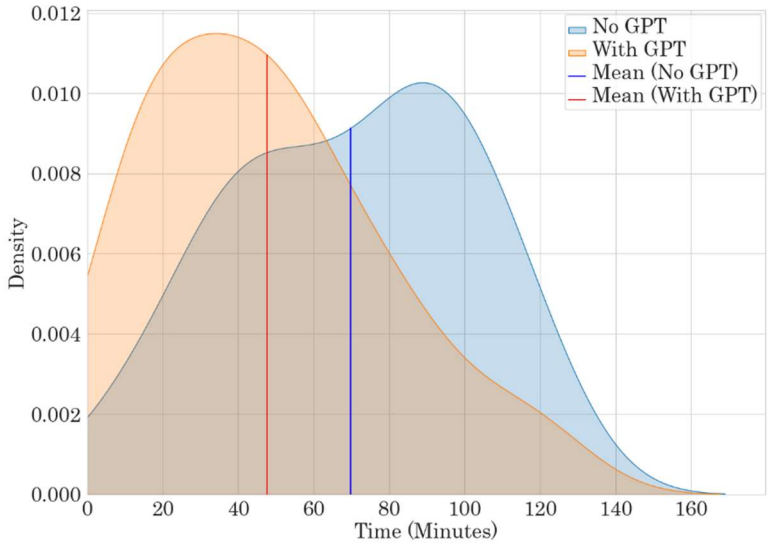


Figure 7: Time Distributions with and Without AI—Employee Handbook

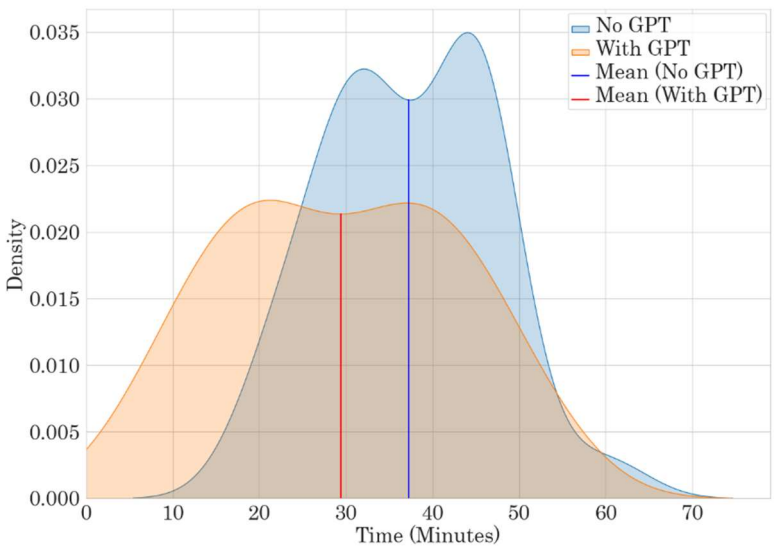
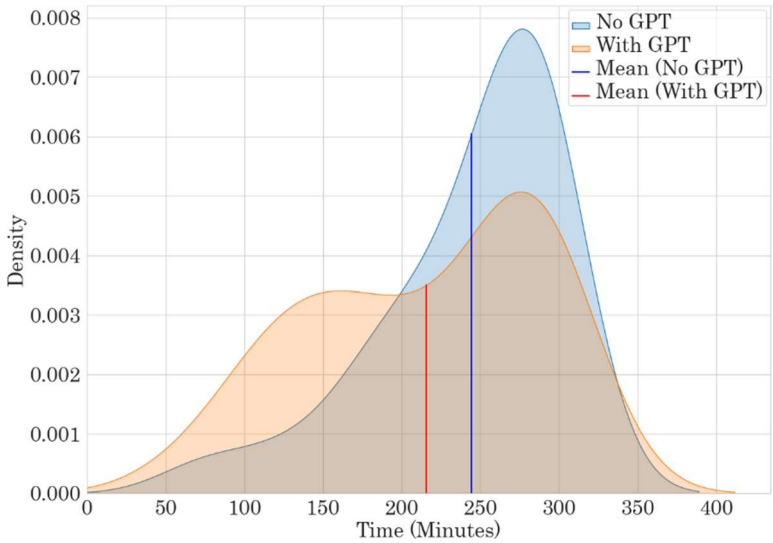


Figure 8: Time Distributions with and Without AI—Client Memo



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 9 through 12 below. Conceptually, if performance is perfectly correlated between tasks, this line should be a forty-five-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.[93] 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.

93. The range of the treatment and control lines on the *x*-axis differ for some of the graphs, because the range of grades awarded to students differed by task, and the tasks available to serve as the treatment and control groups differ depending on the task that is being studied.

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

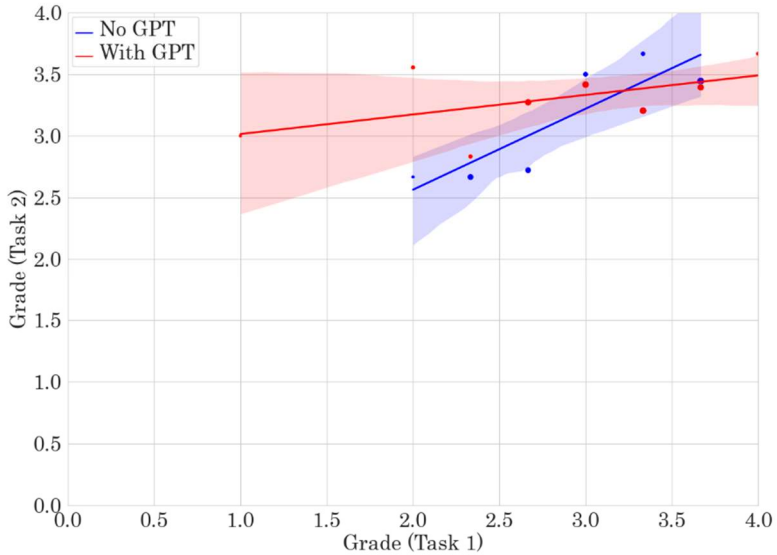


Figure 10: Task 1 vs. Task 2 Grades—Contract Drafting

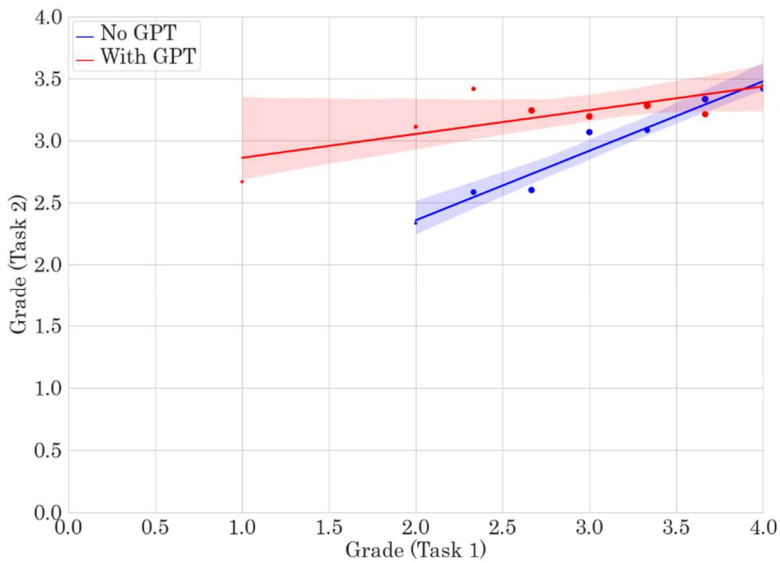


Figure 11: Task 1 vs. Task 2 Grades—Employee Handbook

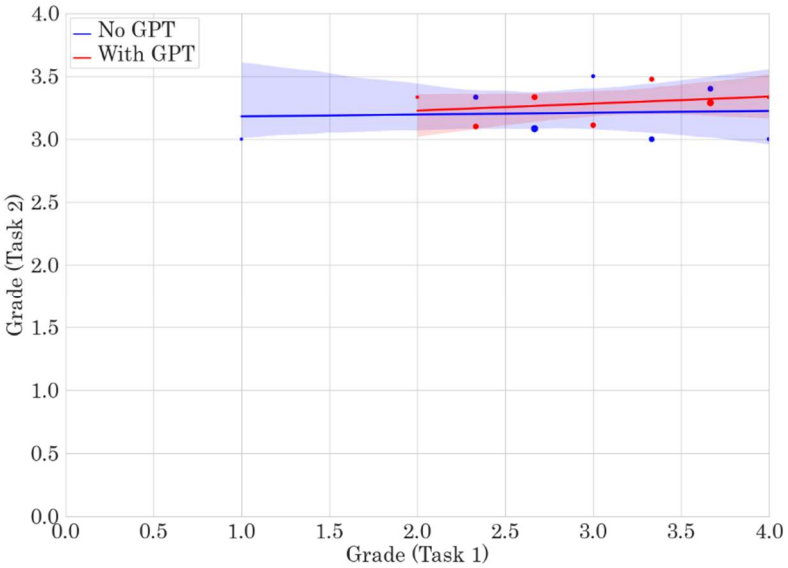
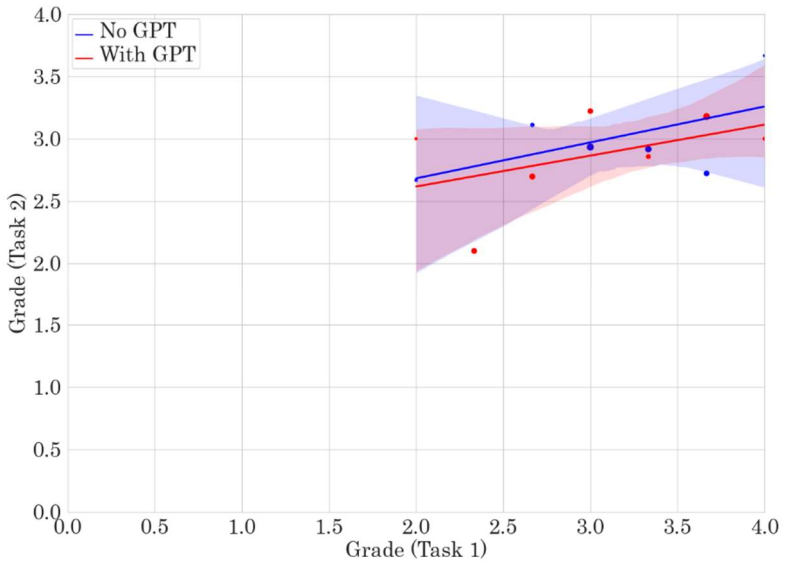


Figure 12: Task 1 vs. Task 2 Grades—Client Memo



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 employee 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 3 below confirms that, for the tasks on which AI assistance was most useful (Complaint Drafting and Contract Drafting) the differences in slopes are large and statistically significant at the ninety-five percent level.

Table 3: Slope of Performance Between Tasks 1 and 2 (Grade)

Task	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 13 through 16 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 ninety-five percent level. Our results therefore suggest that GPT-4 has the potential to reduce time spent on tasks for lawyers of all ability levels.

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

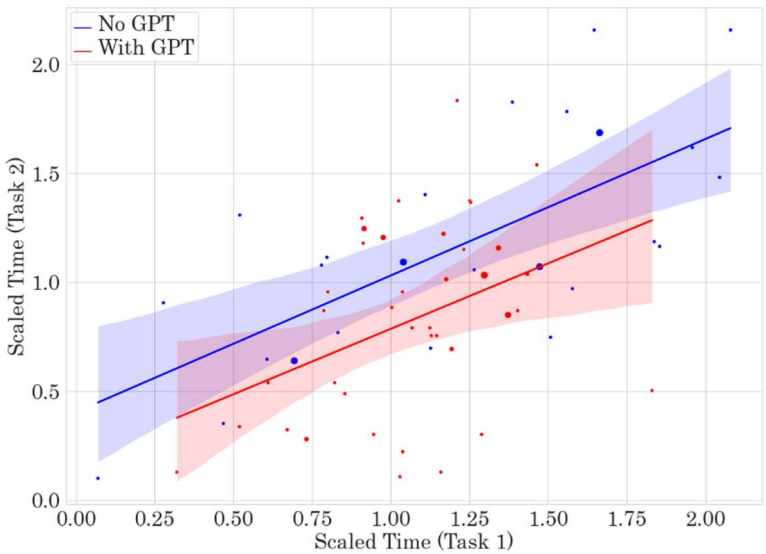


Figure 14: Task 1 vs. Task 2 Time—Contract Drafting

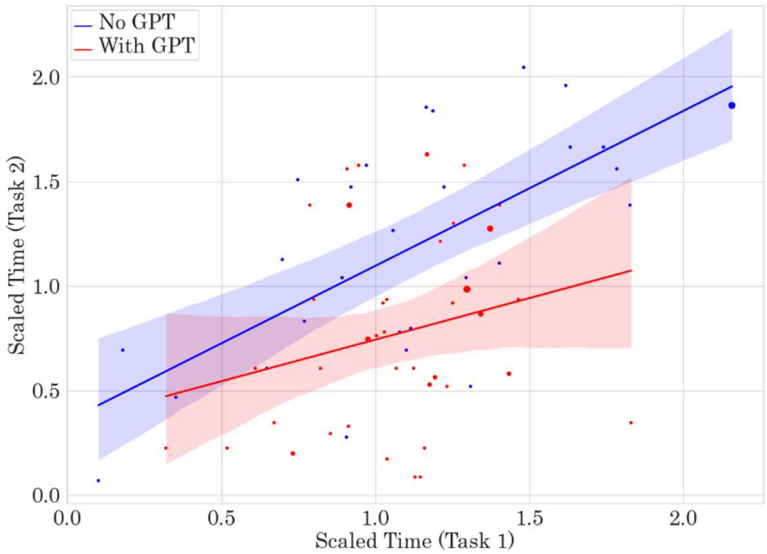


Figure 15: Task 1 vs. Task 2 Time—Employee Handbook

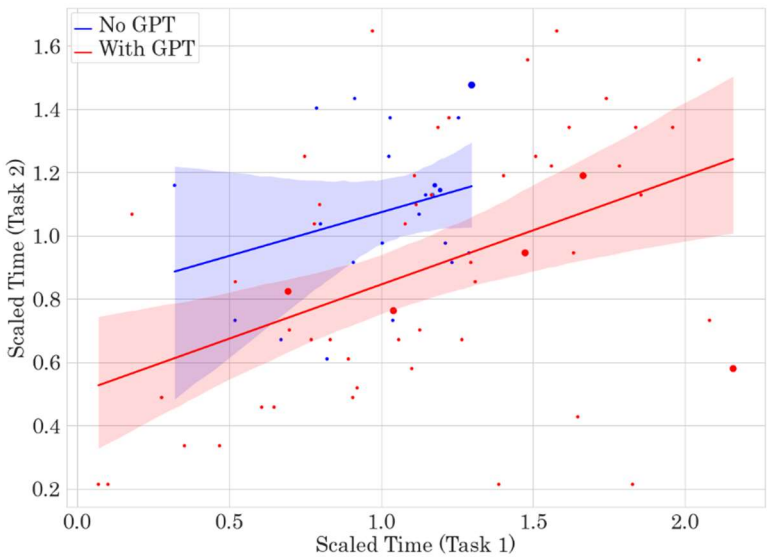


Figure 16: Task 1 vs. Task 2 Time—Client Memo

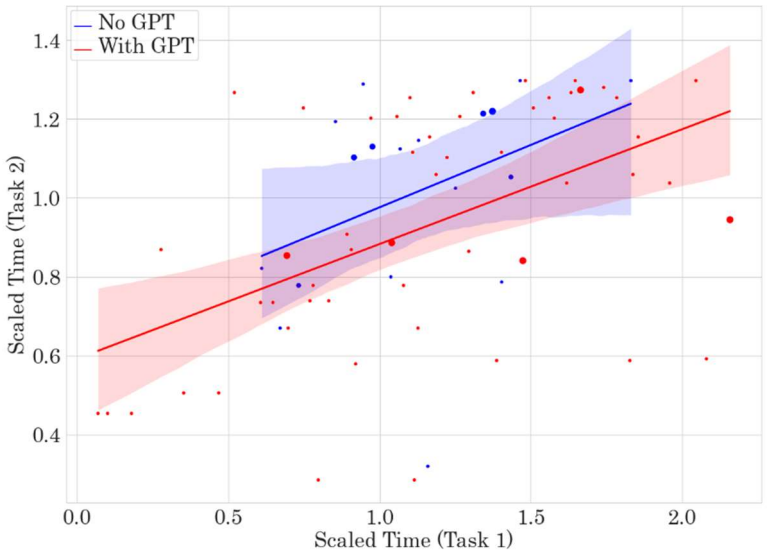


Table 4 reflects these results.

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

Task	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 survey questions asked participants to report their perceptions of the impact GPT-4 access had on the quality of their work and the speed with which they were able to complete tasks. They were also asked to rate their perceived helpfulness of GPT-4 for each individual assignment. In addition, participants were asked about whether they thought their skill using GPT-4 improved over the course of the experiment, whether having access to GPT-4 improved their personal satisfaction with work assignments, and various questions aimed at measuring their interest in using GPT-4 to assist with legal work in the future.

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?

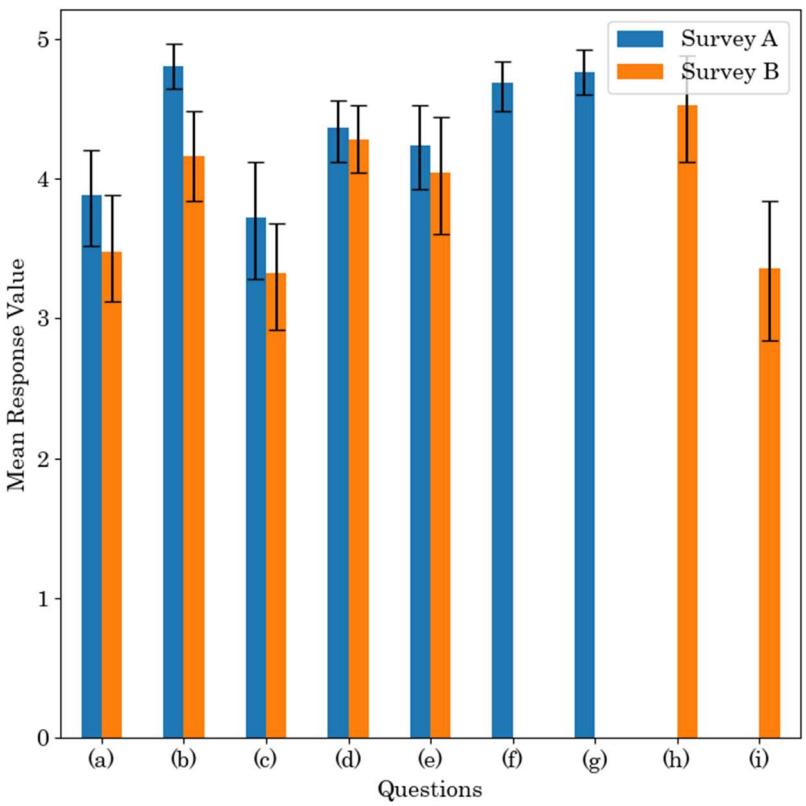
(f) To what extent did you find access to GPT-4 to be helpful for the complaint drafting assignment specifically?

(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?

Figure 17: Survey Results by Question



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). The 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 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. LIMITATIONS, ASSUMPTIONS, AND ROBUSTNESS CHECKS

Although we attempted to design our experiment as cleanly as possible, we inevitably made assumptions or design choices that could potentially limit the robustness or validity of our findings. We describe them here to appropriately frame our results.

First, our experiment had a relatively small sample size, with fifty-nine participants each completing four tasks. Many studies in the literature on human-computer interaction collect far larger samples in order to maximize statistical power; for example, Noy and Zhang gathered a sample of 453 participants for their study of AI's effect on professional writing tasks.[94] The tradeoff is that to keep costs manageable, Noy and Zhang (like many other scholars) recruited participants from a low-cost online survey provider, gave them virtually no training, and had them complete simple, short tasks.

94. Noy & Zhang, *supra* note 70, at 187.

In contrast, we chose to prioritize external validity rather than statistical power. We recruited upper-level law students rather than laypeople, designed realistic lawyering tasks that took an average of 463.5 minutes (7.725 hours) for participants to complete, and provided several hours of training prior to task completion. Scaling the sample size would have ballooned the cost of the study to hundreds of thousands of dollars. However, our choice to focus on external validity comes at the cost of precision, and due to the small sample size, our estimates are generally noisier than we would like.

A second set of limitations relates to the assignment of tasks. All participants completed their tasks in the same order, because we did not want the treatment effect of AI assistance to be confounded with ordering effects—if, for example, we reordered the tasks so that all participants first completed two tasks without AI assistance and then two tasks with assistance, performance improvements might be attributable to increased familiarity with the tasks rather than the AI assistance itself. However, because the tasks were always completed in the same order, it is possible that the ordering interacted with the treatment (AI assistance) in unforeseen ways. For example, it is possible that AI was less useful for the later tasks because AI assistance is more valuable when participants are "fresh." While we do not find this explanation particularly likely, future research could delve further into this issue.

Relatedly, we assigned all participants into one of two groups rather than conducting full randomization. We did this because it made the experiment easier to administer and in order to guarantee that each participant completed two tasks with AI assistance and two tasks without. (When recruiting for the study, we promised to allow all participants the opportunity to contrast their own performance with and without AI.) However, this structure makes it especially important that we conduct effective randomization such that the two groups are identical, so that any differential in their performance can be attributed solely to AI assistance. Otherwise, any differences between treatment and control with respect to each of the tasks could be driven by differences between the groups themselves.

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 first-year first-semester grade point average. We did not collect

other demographic information out of concerns about anonymity. Table 5 provides information about individual characteristics, including means and standard deviations, as well as the difference between the two groups. The differences have p-values of 0.44 for class year and 0.92 for grade point average and do not suggest any statistically significant differences between the two groups.

Table 5: Group A and Group B Individual Characteristics

	Group A	**Group B**	**Diff. (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 Kolmogorov-Smirnov tests to estimate the likelihood that the class years for Group A and Group B, and the grade point averages for Group A and Group B, were drawn from the same distribution. The Kolmogorov-Smirnov statistic for class year was 0.14 (p = 0.95) and for grade point average was 0.10 (p = 1.00), again not suggesting any difference between the two groups.

Although Group A and Group B appear to have been effectively randomized, our grouping methodology has one further implication, specifically for the discussion accompanying Figure 9 through Figure 16. Recall that each of those figures contained a line representing a control (the predicted grade in a specific task without AI assistance based on the participant's grade in another task completed without AI assistance) as well as a treatment (the predicted grade in a specific task *with* AI assistance based on the participant's grade in another task completed without AI assistance). Recall as well that we are looking specifically at the difference in slopes between these two lines. Because each group completed a different set of tasks without AI assistance, we assume that, on average, performance on each task predicts performance on other tasks equally well.

Third, we implicitly make the stable unit treatment value assumption (SUTVA),[95] including the assumption that the performance of the participants in the control group does not differ in light of their assignment to the treatment group on other tasks. It is possible that this assumption is violated to some extent. For example, participants completing tasks without the assistance of AI might subconsciously expect that their performance on unassisted tasks should be worse in comparison to tasks where they have access to AI and therefore might exert less effort on those tasks than they would have outside of an experimental setting, where they were simply completing unassisted AI tasks alone.

There is some evidence to suggest that SUTVA holds against this possibility. Intuitively, students participated in the experiment in part to gauge how much their productivity would improve when given access to AI. They would only receive the benefit of a meaningful comparison if they exerted full effort, giving them some incentive not to shirk. In addition, using time spent completing each task as a proxy for effort, the students spent more time on the tasks without AI assistance, not less, suggesting that any subconscious shirking was marginal.

Fourth, participants in our study were all students or recent graduates of a highly selective law school who expressed interest in participating in a study evaluating the use of AI for legal tasks.[96] As a result, our study participants likely reflect a higher skill level than those of an average law student or recent graduate and may also possess greater technological proficiency and comfort than the average lawyer or law student. Some participants may have had some prior exposure to using generative AI to complete legal tasks.

Fifth, the tasks assigned to study participants were not perfectly representative of tasks that a junior lawyer would face. While we believe that they accurately capture certain key skills, they were simplified in various ways. The client memo did not require independent research, for example, and the contract drafting exercise had material terms specified and had a very

95. *See* Donald B. Rubin, *Which Ifs Have Causal Answers*, 81 J. AM. STAT. ASSOC. 961, 961 (1986) (providing a technical mathematical definition of SUTVA).

96. *See* U.S. NEWS, *supra* note 84 (ranking the University of Minnesota Law School at sixteenth in the U.S. News ranking).

modest scope. Certainly, further study of AI-assistance on more sophisticated lawyering tasks is warranted.

Finally, our instructions regarding the time spent on assignments might also create conditions that would not be perfectly replicated in real world scenarios. Recall that participants were given a maximum amount of time they could spend on each individual assignment but were instructed to submit the assignment when they would feel comfortable turning it in to a supervising attorney. Participants were provided with a flat rate of compensation for their completion of the study, thereby creating an economic incentive to spend as little time as possible on the given tasks. In the real world, under time-based client billing, lawyers have an economic incentive to spend as much time as possible on a task in order to maximize revenue. It may be, therefore, that our findings regarding efficiency will not translate to real world settings. We believe, however, that there are disciplining factors in the real world, including market competition and client pressure, that limit the amount of time a lawyer can reasonably expend on a given task, making our study design a reasonable facsimile of the time pressures faced by a lawyer.

V. IMPLICATIONS

Our results have broad implications for the future of lawyering. Section A of this Part develops these implications by contextualizing our results within the rapidly accelerating development of generative AI technology, both in the legal domain and more generally. This technological development, Section A suggests, means that our results are likely to significantly understate the future potential of AI to impact the work of lawyers. Given this reality, Section B develops the normative implications of our results for lawyers, purchasers of legal services, judges, law schools, and law students. For all of these actors, the bottom line is that generative AI is likely to substantially impact lawyering in the near term, meaning that thoughtful preparation for this eventuality should begin now.

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 certain 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 the benefits of AI for certain tasks, find that access to AI tends to enhance their work satisfaction when it comes to such tasks, 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.[97] Indeed, this trend has already begun, with some lawyers and law firms proactively embracing generative AI.[98] 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 competitors who avoid AI assistance.

The implications of our results become substantially more striking, however, when they are considered in light of the current pace 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.[99]

97. *See supra* Part I (describing the evolution of legal technology tools, including searchable online databases like Westlaw and Lexis).

98. *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/63WN-VX32] (discussing the introduction of an AI chatbot for contract drafting at Allen & Overy); Emily Hinkley, *Mishcon de Reya Is Hiring an 'Engineer' to Explore How Its Lawyers Can Use ChatGPT*, LEGAL CHEEK (Feb. 16, 2023), https://www.legalcheek.com/2023/02/mishcon-de-reya-is-hiring-an-engineer-to-explore-how-its-lawyers-can-use-chatgpt [https://perma.cc/3ZR6-CPLU] (noting Mischon de Reya's LinkedIn job ad posting for a "GPT Legal Prompt Engineer" to work with lawyers).

99. For instance, the firm Casetext recently launched a product known as CoCounsel, which "does document review, deposition preparation, contract analysis, and timeline creation in minutes." *How GenAI Can Enhance Your Legal Work Without Compromising Ethics*, THOMSON REUTERS (Apr. 17, 2024),

Currently available law-specific tools offer lawyers vastly superior capabilities than the general-purpose AIs like GPT-4 that we used in our experiment. These tools improve performance predominantly by marrying generative AIs like GPT-4 with intelligent prompt-engineering and Retrieval Augmented Generation (RAG), which incorporates legal source material. Intelligent prompt engineering bakes into legal tech platforms prompting strategies that are tested and customized to produce useful results for specific types of legal tasks.[100] RAG, the latter approach, allows generative AIs to retrieve relevant content from large legal databases and to use this material to inform its responses.[101] Combined, these two techniques substantially reduce hallucinations and improve the quality of AI-generated output.[102]

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 few hours of online

---

https://legal.thomsonreuters.com/blog/how-genai-can-enhance-your-legal-work-without-compromising-ethics [https://perma.cc/58B6-3L4X]. Within months of CoCounsel's launch, the legal tech giant Thomson Reuters purchased Casetext for $650 Million. *See, e.g.*, REUTERS, *supra* note 23.

100. 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/FXY4-FDU7]. *See generally* Tyler Cowen & Alexander T. Tabarrok, How to Learn and Teach Economics with Large Language Models, Including GPT (George Mason Univ., Working Paper No. 23-18, 2023) (on file with the Minnesota Law Review) (offering advice on engineering prompts). For prompt-engineering advice that is specific to the legal setting, see *AI Tools for Lawyers*, *supra* note 88.

101. *See generally* Patrick Lewis et al., *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*, 33 ADVANCES NEURAL INFO. PROCESSING SYS. 9459 (2020). For discussion of how tools like Casetext use RAG, see *With AI, You Get What You Give*, THOMSON REUTERS (Aug. 2, 2023), https://casetext.com/blog/prompt-engineering-best-ai-output [https://perma.cc/EE6N-F9E9] ("By connecting GPT-4 to a database of 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.").

102. One interesting and untested question is whether and to what extent widespread use of legal AIs might result in homogenized work product and a decrease in the creativity of legal analysis. The graders for our study did not code for work product similarity among those who completed tasks with the assistance of AI, but anecdotally they did not notice "cookie cutter" work product, perhaps because participants edited AI output prior to submission.

training before attempting to use this technology to craft answers to two of the four assignments they completed while participating in the study.[103] 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.[104] 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.[105]

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.[106] 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.[107] Similarly, the capabilities of GPT-4 at the time of this writing (January, 2024) are significantly improved relative to the version that was available to our participants during the experiment in the

103. *See supra* Part II (describing training of study participants).

104. *See supra* Part III (discussing survey results of study participants).

105. *See AI Tools for Lawyers*, *supra* note 88, at 5 ("The quickest route to proficiency with LLMs is the same route to Carnegie Hall: practice, practice, practice.").

106. *See The Great Acceleration: CIO Perspectives On Generative AI*, MIT TECH. REV. (July 18, 2023), https://www.technologyreview.com/2023/07/18/1076423/the-great-acceleration-cio-perspectives-on-generative-ai [https://perma.cc/9YHW-VPVZ] (discussing the revolutionary impact of generative AI).

107. *See, e.g.*, Katz, Bommarito, Gao, & Arredondo *supra* note 2, at 7 (presenting results by legal subject area); *AI Assistance in Legal Analysis*, *supra* note 8 (discussing exam performance); GPT-4's Law School Grades, *supra* note 48 (same). For examples discussing ChatGPT performance, see Lo, *supra* note 63, at 5–6 (presenting GPT-3.5 results); David A. Wood et al., *The ChatGPT Artificial Intelligence Chatbot: How Well Does It Answer Accounting Assessment Questions?*, 38 ISSUES ACCT. EDUC. 81, 82–83 (2023) (presenting ChatGPT results as applied to accounting-specific content); Nori, King, McKinney, Carignan & Horvitz, *supra* note 56 (discussing GPT-4's abilities to solve medical problems); Alejandro Lopez-Lira & Yuehua Tang, Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models 1 (Feb. 28, 2024) (unpublished manuscript) (on file with the Minnesota Law Review) (examining ChatGPT's ability to forecast returns for the stock market with news headlines).

Summer of 2023.[108] For instance, due to model limitations, our participants were required to copy and paste blocks of text from cases or statutes into prompts, and could not use text longer than two to three pages without receiving error messages. Several participants informally complained about this limitation and noted that it slowed them down. With the current model of GPT-4, however, these limitations would not exist because of the AI's significantly expanded context window and its Retrieval into Platform capabilities, which OpenAI introduced in November 2023. LLMs are almost certain to continue to improve in the coming years due to increases in model size and complexity and continuing innovation in the underlying AI architecture.

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 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.[109] 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. [110] Given the relative

108. Briefly, these improvements include significant increases in the model's "context window," Retrieval Augmented Generation capabilities that allow users to upload documents (including cases and statutes), and customizable GPTs that users can build with natural language and publish for others. *See New Models and Developer Products Announced at DevDay*, OPENAI (Nov. 6, 2023), https://openai.com/blog/new-models-and-developer-products-announced-at-devday [https://perma.cc/T8V4-H4XC].

109. *See supra* Part III.

110. *See AI Assistance in Legal Analysis*, *supra* note 8 (reporting "significant variation in how useful AI assistance was to students depending on their baseline performance," with "worst-performing students benefit[ing] 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 Noy & Zhang, *supra* note 70, at 187 (finding that giving college-educated professionals access to AI improved the performance of less-skilled workers more than high-skilled workers); Erik Brynjolfsson, Danielle Li

homogeneity of our participants, however, further study is warranted to determine the extent of AI quality improvement on a broader range of lawyer skill levels.

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.[111] It may also be the

& Lindsey R. Raymond, *Generative AI at Work* 2–4 (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* Falling Asleep at the Wheel, *supra* note 71 (finding that giving professional recruiters access to high-quality AI harmed humans' ability to assess job applications relative to giving them access to lower quality AI tools).

111. *AI Assistance in Legal Analysis*, *supra* note 8, at 5–6 (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).

case, given our participant population, that AI provided the greatest benefit for those tasks participants were least familiar with. While this appears a reasonable hypothesis, we are somewhat skeptical that this distinction has large explanatory power, given that most participants would be unfamiliar with employee handbook drafting, and likely had some exposure to contract drafting.

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, necessitated little if any independent research, could be completed in between one and five hours of time, and 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.[112] 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.[113]

---

112. *See What It Takes to Build an AI Legal Assistant Lawyers Can Rely on*, CASETEXT (May 12, 2023) [hereinafter *What It Takes*], https://casetext.com/blog/building-an-ai-legal-assistant-lawyers-can-trust [https://perma.cc/UR8K-5J78] (discussing prompt engineering and refining CoCounsel's skills).

113. *AI Set to Save Professionals 12 Hours Per Week by 2029*, THOMSON REUTERS, https://www.thomsonreuters.com/en/press-releases/2024/july/ai-set-to-save-professionals-12-hours-per-week-by-2029.html [https://perma.cc/4Q6Y-NX5C] ("Survey respondents predict AI to free up to 12 hours per week within the next five years, with four hours per week saved in the next year alone—the equivalent of adding an additional colleague for every 10 team members. For a U.S. lawyer, this could translate to an additional $100,000 in billable hours.").

Another interesting aspect of our findings is that participants were not only able to accurately assess how useful GPT-4 was at each task, but also that participants reported increased satisfaction when completing tasks with access to GPT-4. With respect to the first finding, this suggests that law firms can be relatively confident that they can trust their lawyers to know when AI will or will not be useful to them in completing a task, rather than having strict controls on AI usage. While the second finding regarding increased satisfaction may at first glance seem a relatively minor point, law firms would do well to take note. In an era where lawyer dissatisfaction and burnout are widespread,[114] a tool that has the potential to increase lawyer well-being, presumably by reducing or eliminating the burden of relatively tedious work, is one that is worth taking seriously.

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 questions, other than to suggest that the assumption that the

114. *See, e.g.*, Jacquelyn Palmer & Linda Ouyang, *Analysis: Survey Finds Lawyer Burnout Rising, Well-Being Falling*, BLOOMBERG LAW (June 28, 2021), https://www.bloomberglaw.com/bloomberglawnews/bloomberg-law-analysis/X15S722S000000?bna_news_filter=bloomberg-law-analysis#jcite [https://perma.cc/64BX-DQES] (reporting that short-term job satisfaction was down while rates of burnout were up, particularly among junior and mid-level associates); Nat'l Task Force on Lawyer Well-Being, *The Path to Lawyer Well-Being: Practical Recommendations for Positive Change*, AM. BAR ASS'N 7 (Aug. 2017), https://www.americanbar.org/content/dam/aba/images/abanews/ThePathToLawyerWellBeingReportRevFINAL.pdf [https://perma.cc/F4U8-JEZZ] (reporting low rates of well-being among early career lawyers and law students).

future will resemble the past is likely tenuous, at best, and that further study is clearly warranted.

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. Lawyers and Law Firms

Our results strongly suggest that lawyers and law firms should be proactively exploring how best to incorporate generative AI tools into their practice. Of course, many law firms are already 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 tools for lawyers, into its practice.[115] Numerous other large law firms have also embraced this tool in recent months, though many have been reluctant to publicly acknowledge this.[116] Other large global law firms—including Allen & Overy—have incorporated a competing generative AI tool, Harvey, into their practice.[117] Still other firms have taken a different approach, hiring their own AI

115. Press Release, DLA Piper to Utilize CoCounsel, The Groundbreaking AI Legal Assistant Powered by OpenAI Technology (Mar. 15, 2023), https://www.dlapiper.com/en-us/news/2023/03/dla-piper-to-utilize-cocounsel-the-groundbreaking-ai-legal-assistant-powered-by-openai-technology [https://perma.cc/8XDM-BALJ].

116. *See* Press Release, Casetext, Top Global Law Firm DLA Piper Announces Addition of CoCounsel to Enhance Practice and Client Services (Mar. 23, 2023), https://casetext.com/blog/law-firm-dla-piper-announces-casetext-cocounsel [https://perma.cc/YG8B-W2SW] (listing other firms that have adopted CoCounsel).

117. Charlotte Johnstone, *MacFarlanes Joins List of Firms Adopting Harvey AI*, LAW.COM (Sept. 21, 2023), https://www.law.com/international-edition/2023/09/21/macfarlanes-joins-list-of-firms-adopting-harvey-ai [https://perma.cc/RHS6-ETVE].

experts to develop proprietary and firm-specific generative AIs that are not available to competitors.[118]

Although this trend is already evident in large law firms, 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 New York lawyer who relied on ChatGPT to author a brief without double-checking the resulting output.[119] 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.[120] 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.[121]

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. 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

118. *See* Lance Eliot, *Prestigious Symposium on AI Lawyering Reveals Keen Insights Including the Ardent Debate on Whether to Use Generative AI in Law School Education*, FORBES (Oct. 17, 2023), 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 [https://perma.cc/D2M6-NSNW].

119. *See* Benjamin Weiser, *Here's What Happens when Your Lawyer Uses ChatGPT*, N.Y. TIMES (May 27, 2023), https://www.nytimes.com/2023/05/27/nyregion/avianca-airline-lawsuit-chatgpt.html [https://perma.cc/F62R-484L].

120. *Id.*

121. *See id.* (""The discussion now among the bar is how to avoid exactly what this case describes,' Mr. Gillers said. 'You cannot just take the output and cut and paste it into your court filings.'").

made by AIs; (ii) experimenting with different prompting strategies when using general purpose AIs, including few-shot and grounded prompting;[122] (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 assurances of confidentiality.[123] AI will be more useful in some practice areas than others, and lawyers should take the time to become familiar with it to use it most effectively.

2. Legal Clients

The potential for generative AI to significantly improve the efficiency of legal work should be welcome news to many clients. 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.[124] 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.[125] Some lawyers may be inclined to accomplish this simply be resisting incorporating generative AI into their

122. Few-shot prompting involves providing the AI model with examples of good responses that it can use to shape its response. *See* Michael Bullwinkle & Eric Urban, *Prompt Engineering Techniques*, MICROSOFT LEARN (Sept. 5, 2024), https://learn.microsoft.com/en-us/azure/ai-services/openai/concepts/advanced-prompt-engineering?pivots=programming-language-chat-completions [https://perma.cc/2CYC-4VCG] (describing few-shot as providing examples "to give additional context to the model"). Grounded prompting involves providing the AI model with relevant sources. *Id.* (describing giving a "model data to draw its responses from" as grounding).

123. *See AI Tools for Lawyers*, *supra* note 88, at 20–21 (discussing confidentiality issues).

124. *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) (discussing the need for standards to avoid conflicts and adverse representation).

125. *See* Lisa G. Lerman, *A Double Standard for Lawyer Dishonesty: Billing Fraud Versus Misappropriation*, 34 HOFSTRA L. REV. 847, 848–49 (2006) (discussing lawyer dishonesty regarding finances and unequal disciplinary enforcement).

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 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.[126] 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 like high-stakes litigation or corporate mergers and acquisitions where outcomes matter much more than the size of the legal bills.

---

126. *See* John Armour & Mari Sako, *AI-Enabled Business Models in Legal Services: from Traditional Law Firms to Next-Generation Law Companies*, 7 J. PROS. & ORG. 27, 27 (2020) (evaluating how the adoption of AI in the legal services will impact the structure of law firms generally, including the extent to which clients will choose to develop new forms of expertise internally).

3. Judges

In the wake of several recent high-profile stories of lawyers using ChatGPT to produce legal filings with significant errors,[127] many judges have adopted formal policies regarding the use of generative AI by lawyers practicing before them. An increasing number of judges, for instance, require lawyers to disclose whether they used generative AI to help them write legal filings.[128] Other judges go further, requiring lawyers to specially certify the accuracy of any filings for which generative AI has been used.[129] And several judges have even prohibited lawyers that practice before them from using any generative AI to assist them with writing legal filings.[130]

In our view, our results suggest that such aggressive attempts to limit or complicate lawyers' use of generative AI are misguided.[131] Generative AI has the capacity to allow lawyers to

127. *See, e.g.*, Ella Lee, *Michael Cohen Gave Lawyer Fraudulent Case Citations Generated by AI*, THE HILL (Dec. 29, 2023), https://thehill.com/regulation/court-battles/4381736-michael-cohen-gave-lawyer-fraudulent-case-citations-generated-by-ai [https://perma.cc/95UW-ET66] (revealing in court documents that Michael Cohen had given his attorney fake legal cases); Larry Neumeister, *Lawyers Submitted Bogus Case Law Created by ChatGPT. A Judge Fined Them $5,000*, ASSOCIATED PRESS (July 22, 2023), https://apnews.com/article/artificial-intelligence-chatgpt-fake-case-lawyers-d6ae9fa79d0542db9e1455397aef381c [https://perma.cc/8CUH-R2P9] (reporting a federal judge fining two lawyers for fictitious legal research).

128. *See* Odia Kagan, *Federal Judges Start Cracking Down on the Use of Artificial Intelligence in Court Filings*, FOX ROTHSCHILD (Dec. 11, 2023), https://dataprivacy.foxrothschild.com/2023/12/articles/artificial-intelligence/federal-judges-start-cracking-down-on-the-use-of-artificial-intelligence-in-court-filings [https://perma.cc/MZ7M-WSD4] (discussing proposed and existing AI disclosure rules in federal courts).

129. *See* Shweta Watwe, *Judges Reflect on GenAI Use One Year After ChatGPT's Debut*, BLOOMBERG LAW (Nov. 28, 2023), https://news.bloomberglaw.com/litigation/judges-reflect-on-genai-use-one-year-after-chatgpts-debut [https://perma.cc/L34B-NFYR] (highlighting additional certifications required by some judges when AI is used).

130. *See* Megan Cerullo, *Texas Judge Bans Filings Solely Created by AI After ChatGPT Made up Cases*, CBS NEWS (June 2, 2023), https://www.cbsnews.com/news/texas-judge-bans-chatgpt-court-filing [https://perma.cc/YEJ4-QEWB] (describing a Texas judge's rules on AI usage).

131. It is certainly possible that the development of generative AI will impact judges' own drafting of their judicial opinions as well. *See* Richard M. Re, *Artificial Authorship and Judicial Opinions*, 92 GEO. WASH. L. REV. (forthcoming 2024) (manuscript at 11) (on file with the Minnesota Law Review) (speculating that generative AI may significantly impact the quantity and quality of judicial opinions).

better serve their clients by producing work product more efficiently, thus reducing barriers to justice.[132] Imposing special restrictions on lawyers' use of this technology not only tends to counteract this salutary effect, but also to stigmatize the use of generative AI more generally. While lawyers can of course use this technology irresponsibly to produce fabricated citations or source material, the possibility of such malpractice is hardly limited to generative AI. To the contrary, new technologies ranging from e-discovery platforms to searchable legal databases create their own distinct risks of malpractice. These risks, as well as virtually all other risks of attorney misconduct, have historically been regulated by general rules of professional conduct that are not tied to specific legal technologies or subject areas.[133] Just as these general rules of professional responsibility have been flexible enough to deter and penalize past misuses of legal technology, so too are they flexible enough to deter and penalize the inappropriate use of generative AI by lawyers today.[134]

4. Law Schools and Law Students

Given the potential of generative AI to impact the practice of law, it is no wonder that law schools across the country are grappling with how to incorporate AI into their curricula.[135]

132. *See* Roberts, *supra* note 1, at 6 ("Rule 1 of the Federal Rules of Civil Procedure directs the parties and the courts to seek the 'just, speedy, and inexpensive' resolution of cases [and many] AI applications indisputably assist the judicial system in advancing those goals." (quoting FED. R. CIV. P. 1)).

133. *See generally* Jon J. Lee, *A New Approach to Attorney Regulation*, 65 B.C. L. REV. 1625 (2024) (noting that rules governing attorney misconduct have historically been general in nature).

134. Indeed, the infamous New York lawyer who used ChatGPT to produce fabricated citations was sanctioned under Rule 11 by the presiding judge. *See* Sara Merken, *New York Lawyers Sanctioned for Using Fake ChatGPT Cases in Legal Brief*, REUTERS (June 26, 2023), https://www.reuters.com/legal/new-york-lawyers-sanctioned-using-fake-chatgpt-cases-legal-brief-2023-06-22 [https://perma.cc/T588-53Q3] (describing the sanctions a lawyer received for using ChatGPT to draft a legal brief).

135. *See, e.g.*, Joseph Landau & Ron Lazebnik, *Law Schools Must Embrace AI*, NAT'L L. J. (July 10, 2023), https://www.law.com/nationallawjournal/2023/07/10/needs-edit-law-schools-must-embrace-ai [https://perma.cc/X4BU-YNAC]; Kristen Baginski & Celeste Pometto DiNicola, *AI Goes to Law School*, LEXISNEXIS (Dec. 12, 2023), https://www.lexisnexis.com/community/insights/legal/b/thought-leadership/posts/ai-goes-to-law-school [https://perma.cc/C2F7-Q8RG] ("Law students will soon be actual lawyers so there will be an expectation that those students can use relevant legal AI tools to be efficient and effective

Historically, shifts in legal technology have only had a limited effect on legal training, which 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.[136] Although recent decades have seen important adaptations to this approach—from more inclusive Socratic questioning,[137] to an increased focus on statutory interpretation,[138] to increased opportunities for formative feedback[139]—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.[140] Not even technological change as significant as generative AI is likely to 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

---

practitioners. This means teaching students how to use AI to support critical thinking and evaluation, collaboration and communication, assessment and feedback.").

136. *See* L. Danielle Tully, *What Law Schools Should Leave Behind*, 2022 UTAH L. REV. 837, 837 (2022) (lamenting the lack of change in legal education in recent decades notwithstanding common calls for fundamental reform); 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, 118 (2022) ("For more than 150 years, legal education has largely followed the course charted by Christopher Columbus Langdell when he became dean of Harvard Law School in 1870.").

137. Jamie R. Abrams, *Legal Education's Curricular Tipping Point Toward Inclusive Socratic Teaching*, 49 HOFSTRA L. REV. 897, 898 (2021) (advocating for an inclusive form of Socratic instruction that is "student-centered, skills-centered, client-centered, and community-centered").

138. *See generally* 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) (exploring how the increasingly common practice of Legislation and Regulation during law students' 1L year impacts the upper-level law school curriculum).

139. *See generally* Daniel Schwarcz & Dion Farganis, *The Impact of Individualized Feedback on Law Student Performance,* 67 J. LEGAL EDUC. 139 (2017) (reporting that providing formative feedback to first-year law students on midterm exams improved students' performance in their other first-year classes).

140. *See, e.g.*, Cass R. Sunstein, *On Analogical Reasoning*, 106 HARV. L. REV. 741, 742 (1993) (exploring the distinctive nature of legal reasoning).

answers, confirm the accuracy of arguments and sources, organize issues clearly, and assess the strength of alternative arguments.[141]

For these reasons, law schools should consider substantially limiting the use of generative AI in certain law school classes, particularly classic first-year classes like Contracts and Torts. Because generative AI does not impact the nature of legal reasoning, it need not alter the way that such reasoning is taught by instructors or demonstrated by students, particularly introductory law students.[142] 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.[143]

However, our results suggest that accomplishing this goal requires law schools to proactively limit access to generative AI during student assessments. 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.[144] 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 current technology, law professors who intend to limit access to AI must place hard technological limits or employ aggressive proctoring.[145] Relying instead on honor codes is simply

141. *See AI Tools for Lawyers*, *supra* note 88, at 4.

142. *See id.*

143. *See* Erin McCauliff, *The Calculator in the Elementary Classroom: Making a Useful Tool out of an Ineffective Crutch*, CONCEPT, Apr. 4, 2004, at 1, 2.

144. *See AI Assistance in Legal Analysis*, *supra* note 8, at 5.

145. *See* Julianne Hill, *Profs Trade Notes as Law Schools Write Generative AI Policies*, ABA J. (Jan. 2, 2024), https://www.abajournal.com/web/article/law-profs-trade-notes-as-law-school-write-generative-ai-policies [https://perma.cc/

impractical given the current power of widely accessible generative AI tools.[146] 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.[147] All of this means that cheating among a non-trivial number of students is inevitable when 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, and thus eventually deciding to cheat themselves.[148]

While law schools might restrict student access to generative AI tools in some classes, we believe 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.[149] 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.

2JHV-N5WP] (describing different law schools' efforts to revise their academic integrity codes in response to generative artificial intelligence).

146. *Id.* (discussing the increasing availability of generative AI tools).

147. *See, e.g.*, Zhengyuan Jiang, Jinghuai Zhang & Neil Zhenqiang Gong, Evading Watermark Based Detection of AI-Generated Content (Nov. 21, 2023) (unpublished manuscript) (on file with the Minnesota Law Review) (discussing the difficulty of detecting AI-generated text even if sophisticated technological techniques for "watermarking" such text is attempted).

148. *See* Daniel Houser, Stefan Vetter & Joachim Winter, *Fairness and Cheating*, 56 EUROPEAN ECON. REV. 1645, 1645 (2012) (reporting the results of an experiment suggesting that "individuals who believe they were treated unfairly in an interaction with another person are more likely to cheat in a subsequent unrelated game"); Scott S. Wiltermuth, *Cheating More when the Spoils Are Split*, 115 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 157, 157 (2011) ("We cheat because we think others are cheating . . . .").

149. *See supra* Part III (describing survey results of participants).

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.[150] 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.[151]

CONCLUSION

We conducted the first randomized controlled trial to evaluate LLM assistance with basic lawyering tasks. We found small and variable improvements to the quality of work product but large and consistent improvements to speed. Moreover, we found that when AI provides a boost to quality at all, the boost to quality (but not speed) inversely correlates with baseline performance, with a substantial improvement for the worst performers but no improvement for the best. Finally, we found 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 improving the performance of lower-skilled attorneys.

---

150. *See* Jason G. Dykstra, *Beyond the "Practice Ready" Buzz: Sifting Through the Disruption of the Legal Industry to Divine the Skills Needed by New Attorneys*, 11 DREXEL L. REV. 149, 214 (2018) ("[S]tudents must emerge from law school both ready for practice and prepared to immediately generate revenue, whether they ply their practice-ready skills as contract attorneys, associates, in-house counsel, or solo practitioners.").

151. *See AI Tools for Lawyers*, *supra* note 88, at 4 (arguing that many of the tools traditionally required to be an effective lawyer are also useful in effectively using AI to help produce legal work product).

APPENDIX

A. 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.[152] 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.[153] Training was split into three sub-areas. The first covered general principles on using AI effectively in legal research and writing.[154] Among other things, it provided participants with an overview of basic prompting techniques that prior research had shown to be effective in legal analysis, such as supplying the AI with relevant legal rules or source materials within prompts.[155] Second, the training covered basic techniques for

152. This training drew heavily on previous work by two of us. *See generally AI Tools for Lawyers*, *supra* note 88.

153. 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.

154. 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 an initial assessment of a project by asking it to produce an outline, identify key issues, or produce a first draft (in the case of shorter assignments); (iv) chunk up the elements of an outline, the issues, and application of rules into bite-sized bits, and ask AI to analyze each bit, adjusting the level of generality based on the problem and quality of answers; (v) provide AI with all the key details that a person would need to accomplish the prior step; (vi) iterate by providing additional details that you may have left out, such as by asking AI to alter elements that do not look good, or asking AI to elaborate on elements that do look promising; (vii) provide AI with relevant source materials, including cases, statutes, contract parameters, etc.; and (viii) do not rely on AI to conduct specific legal research or identify specific legal source material unless you confirm the veracity of that material.

155. *See AI Assistance in Legal Analysis*, *supra* note 8, at 22 (discussing prompting strategies for AI usage). For a review of the computer science literature on these prompting strategies, see for example *Prompt Engineering*,

using AI effectively in litigation-oriented settings, covering topics such as using AI to summarize and apply primary sources like caselaw and statutes.[156] 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 the format, style, and structure of sample transactional materials and to help identify alternative terms, unanticipated risks, and ambiguities in initial drafts.[157]

B. ASSIGNMENTS

We selected the four assignments that we gave to participants to be representative of the types of tasks that junior lawyers perform. These assignments were as follows:

OPENAI PLATFORM, https://platform.openai.com/docs/guides/prompt-engineering [https://perma.cc/J2MV-MJ7D] (explaining OpenAI's recommended prompting strategies); Alan D. Thompson, *Microsoft Bing Chat (Sydney/GPT-4),* LIFE ARCHITECT (Feb. 22, 2023), https://lifearchitect.ai/bing-chat [https://perma.cc/H4NR-ATD8] (describing Microsoft Bing's implementation of OpenAI and its prompts); Cowen, *supra* note 100 (providing tips and guidance for the usage of AI in economics problems). *See generally What It Takes*, *supra* note 112 (detailing the work done behind the scenes in the creation of an AI legal assistant); Jason Wei et al., *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models*, 35 ADVANCES NEURAL INFO. PROCESSING SYS. 24824 (2022) (discussing the usage of chain of thought models in AI learning); Tom B. Brown et al., *Language Models Are Few-Shot Learners*, 33 ADVANCES NEURAL INFO. PROCESSING SYS. 1877 (2020) (testing AI models in few-shot settings); 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 the Minnesota Law Review) (exploring strategies to reduce AI hallucinations and generate better results).

156. 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 GPT-4 (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.

157. 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 the terms are provided; (iii) identify potential risks to address and ambiguities in deal terms; (iv) help issue-spot potential additional terms to add to an agreement; and (v) help further develop/specify terms, or identify alternative ways of drafting that can favor one particular side in the transaction.

(1) Legal Memo Assignment

Chris Smith was known in his community as an uncannily talented grill master, in part because of his excellent homemade barbeque sauce, a family recipe. After years of friends suggesting that he make money on his family recipe, Smith decided to market it commercially.

Smith contracted with ABC Food Company to design a hot and spicy version of his sauce for commercial sales. ABC will also manufacture, market and distribute the sauce. Fran Jones, a developer at ABC, was put in charge of the project. Jones wants to design the sauce using serrano peppers for added spiciness, as she believes the serrano is perfect for making Smith's recipe spicier without taking away from the original flavor. However, Jones is concerned about using serrano peppers because she knows, from personal experience, that some people are allergic to it. Jones has seen each of two friends break into rashes upon eating the peppers. In addition, a study commissioned by the American Hyper Allergy Association of America! (AHAAA!) has projected that up to 1% of Americans may have a propensity for allergic reaction to the pepper. This reaction will likely take the form of a rash in most of the pepper-sensitive population, but the reaction could involve an acute and therefore potentially life-threatening increase in blood pressure in a subset of that population.

As a first-year associate lawyer for ABC, it falls to you to determine the legal implications of using the serrano pepper in Smith's barbeque sauce. There are a number of common law and statutory issues presented. ABC is aware that both federal food and drug law, and statutory enactments in various states including ABC's home state of Ohio, may preempt or at least supplement common law. But you have been asked for now to examine only the question whether a warning is required to avoid strict liability under Restatement (Third) of Torts: Products Liability § 2, cmt. k, and Restatement (Second) of Torts § 402A, cmts. h, i, and especially j.

One of your colleagues, a third-year associate, has already done extensive research into the case law on this matter (the four most illustrative cases she found are included below). ABC for now does not want additional research. Your assignment is to review the four cases your colleague has already found, and then to write an objective, predictive memo for ABC on the specific topic of whether ABC needs to put a warning on the barbeque

sauce label if it wants to include serrano peppers as an ingredient in the Smith sauces, and also wants to avoid a risk of strict liability for failure to warn.

ABC and Smith are located in Ohio, and any initial distribution of Smith's sauce will likely take place in Ohio for test-marketing purposes.

Relevant cases

Crislip v. TCH Liquidating Co., 52 Ohio St. 3d 251 (1990)
Mills v. Giant of Maryland, LLC, 508 F.3d 11 (D.C. Cir. 2007)
Livingston v. Marie Callender's, Inc., 72 Cal. App. 4th 830 (Cal. Ct. App. 1999)
Adelman-Tremblay v. Jewel Companies, Inc., 859 F.2d 517 (7th Cir. 1988)

(2) Contract Drafting Assignment

Jill Jackson wants to employ Mary Monte to paint four rooms (living room, dining room, kitchen, and downstairs bathroom) in her home. She is willing to pay for all materials, including paint, brushes, etc. immediately upon presentation of receipts and pay $3,000 total when the job is completed. She anticipates that the job will not require any primer, but that all surfaces will need two coats. The color of all trim will be Sherwin Williams "bright white" and the color of all walls will be Sherwin Williams "shadow gray." Jill wants the work done no later than 6 weeks from the date of the contract, because only a week later she will be hosting her son's graduation party. She is willing to pay 20% of the $3,000 upon execution, and the remainder when the work is done to her satisfaction. Please draft a contract favorable to the homeowner, which is in plain English. Both parties are located in the state of Minnesota. The contract should be no more than two pages single-spaced (12-point type, 1" margins).

(3) Employee Handbook Assignment

Sergio and Stella are software developers based in Minneapolis, MN. They started Code Castle LLC two years ago and have run it themselves since then. Now, with more work than they can handle, Sergio and Stella have hired three employees and expect to hire more next year. Starting next week, Maria and Mo will join them as full-time developers and Mattias will

be the office manager. Code Castle purchased an "off-the-shelf" employee handbook that they've been using, but they have realized it is missing some information they believe is important to cover. They have hired your firm to help them revise the handbook to include various topics not included in the basic handbook they purchased.

One topic they would like added to the handbook is employee breastfeeding accommodations. This particular topic arose because Maria has a 3-month-old baby and Sergio and Stella want to be sure they understand what they need to do to accommodate Maria pumping breastmilk while at work.

Your supervising attorney has asked you to draft a section to add to the employee handbook that explains an employee's rights under applicable law to pump breastmilk while at work. Please research relevant state and federal law and provide a draft of the requested section. Please make sure the section is no longer than one page.

(4) Complaint Drafting Assignment

Unlike the first three assignments, elements of the complaint-drafting assignment are occasionally re-used by the instructor who designed it. As such, that instructor has requested that we not publicly disseminate the full content of the assignment. In brief, however, this assignment requires students to draft a legal complaint for a federal court based on a two-page memo from a client describing how his restaurant and bar experienced unfair treatment by the local police and other authorities. The memo includes numerous details, some of which are quite relevant to establishing a potential civil cause of action, and others of which are either less relevant or completely irrelevant. The memo also specifies four particular legal theories that the complaint could assert and provides students with the underlying elements of these causes of action. It does not contain any details regarding the appropriate form or content of a complaint under the Federal Rules of Civil Procedure.

C. 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

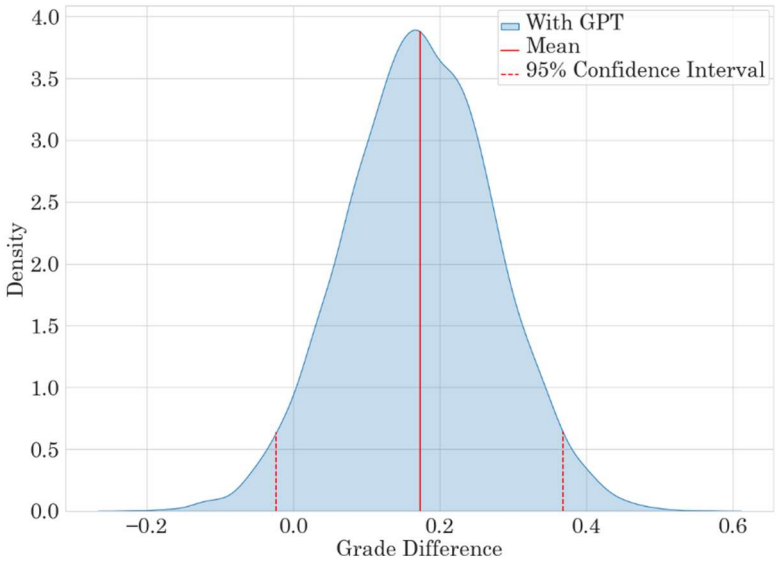


Figure 19: Difference in Grade with Access to AI—Contract Drafting

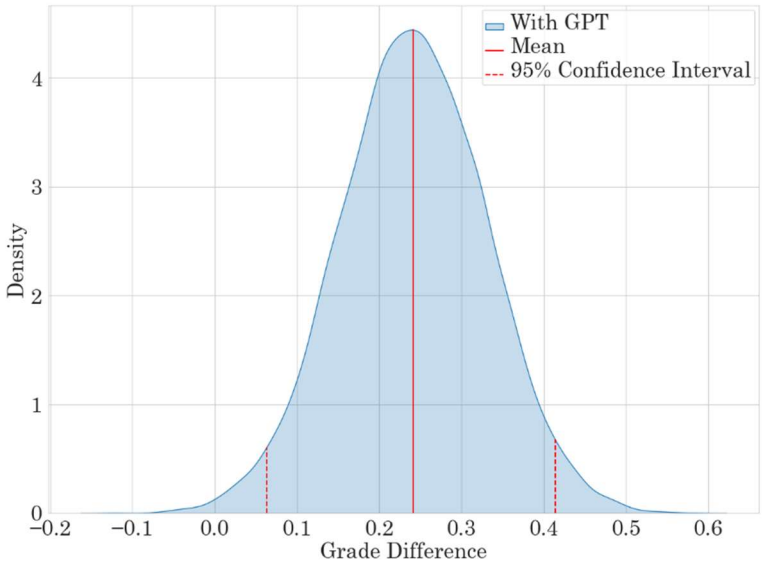


Figure 20: Difference in Grade with Access to AI—Employee Handbook

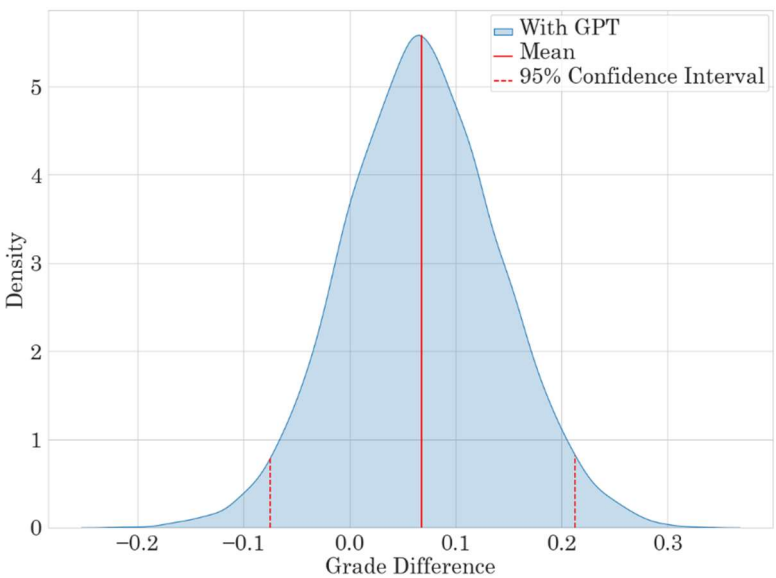


Figure 21: Difference in Grade with Access to AI—Client Memo

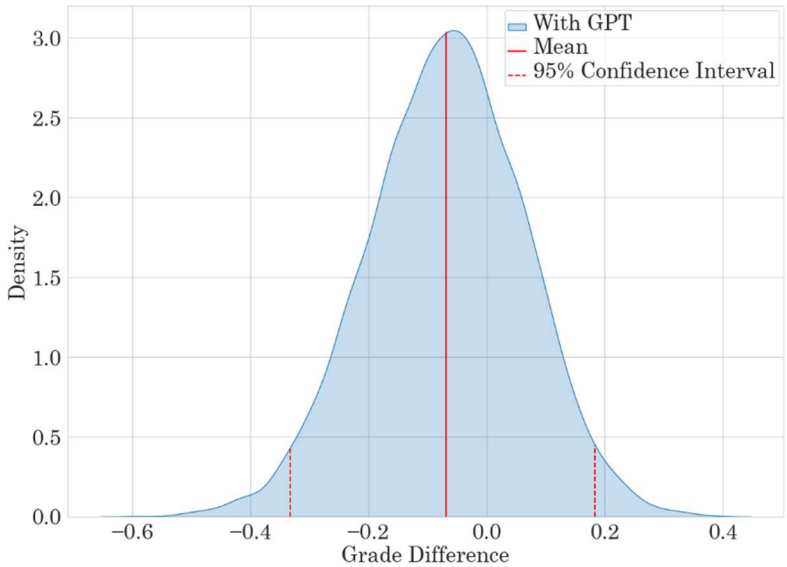


Figure 22: Difference in Time Taken with Access to AI—Complaint Drafting

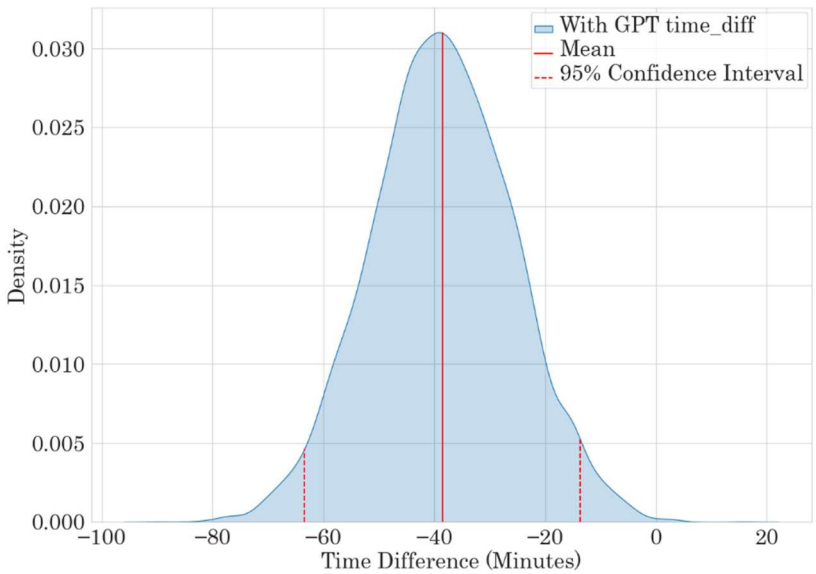


Figure 23: Difference in Time Taken with Access to AI—Contract Drafting

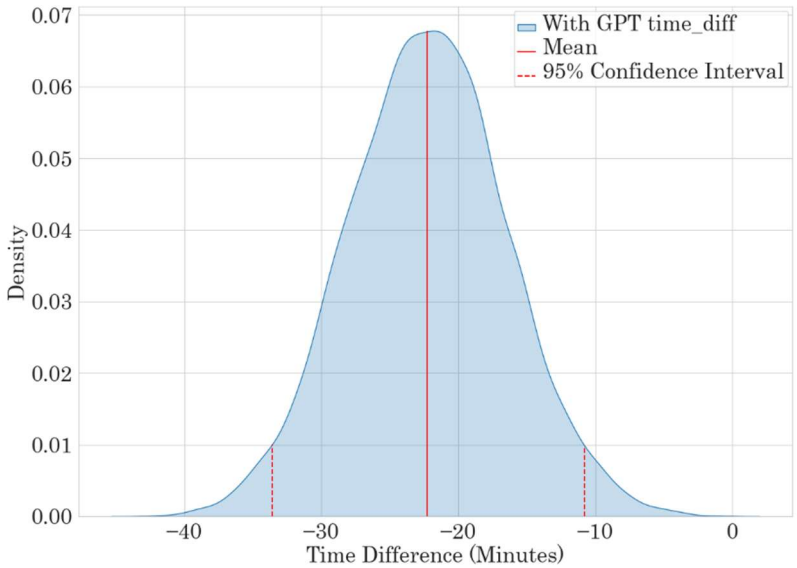


Figure 24: Difference in Time Taken with Access to AI—Employee Handbook

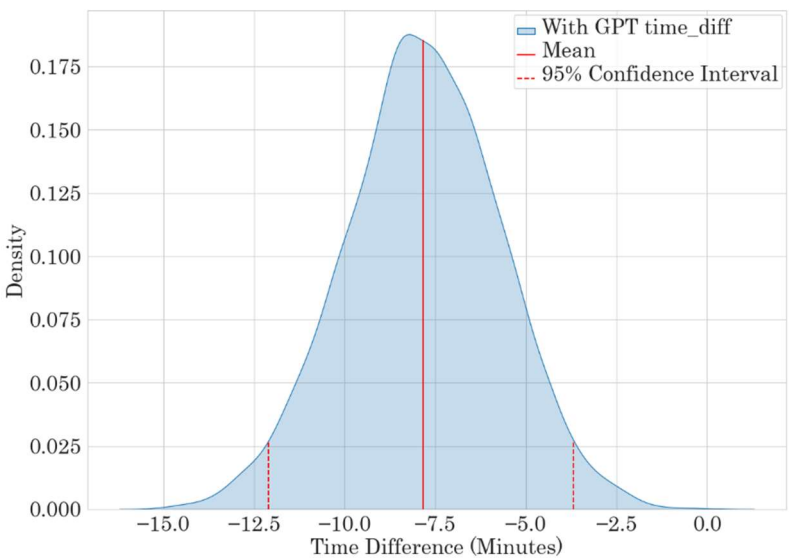


Figure 25: Difference in Time Taken with Access to AI—Client Memo