DRAFT, May 19, 2025

THINKING LIKE A LAWYER IN THE AGE OF GENERATIVE AI: COGNITIVE LIMITS ON AI ADOPTION AMONG LAWYERS

Daniel Schwarcz, [*] ***Debarati Das,*** [**] ***Dongyeop Kang***[***] ***& Brett McDonnell***[***]

Abstract

As of mid-2025, there is robust evidence that generative AI possesses the technological capability to significantly reshape legal practice. Yet legal markets and doctrines have, to date, remained largely unchanged. This gap is often attributed to familiar lags in technology adoption and a range of socio-legal and institutional constraints. This Essay offers a complementary explanation: lawyers face distinctive and under-appreciated challenges to deeply understanding AI-generated outputs to complex or unfamiliar legal problems. These difficulties, which have nothing to do with "hallucinated" sources, hamper lawyers' ability to evaluate the quality of AI assistance and to perform closely related tasks—such as engaging with clients or judges, tailoring arguments to new contexts, or synthesizing insights across legal issues. Although these limitations may diminish as the profession adapts, they may also reflect more fundamental features of human legal reasoning, particularly among junior and less experienced attorneys. If so, these dynamics are likely to influence AI's role in legal practice for the foreseeable future, while also offering critical insights into how lawyers ought to engage with AI tools.

[*] Fredrikson & Byron Professor of Law, University of Minnesota Law School. Schwarcz@umn.edu.

[**] PhD Student, Computer Science and Engineering, University of Minnesota.

[***] Assistant Professor, Computer Science and Engineering, University of Minnesota.

[***] Dorsey & Whitney Professor of Law, University of Minnesota Law School.

Table of Contents

INTRODUCTION **2**
I. THE LEGAL AI GAP **5**
A. AI'S CURRENT POWER TO TRANSFORM LEGAL PRACTICE 5
B. LIMITED EVIDENCE OF AI-DRIVEN LEGAL MARKET TRANSFORMATIONS 10
II. THE DIFFICULTY HUMAN LAWYERS FACE EVALUATING THE QUALITY OF AI OUTPUT **12**
A. THE DIFFICULTY JUNIOR LAWYERS FACE EVALUATING THE QUALITY OF AI OUTPUT ... 12
B. EXPLAINING WHY HUMAN LAWYERS MAY HAVE DIFFICULTY EVALUATING THE QUALITY OF LEGAL WORK PRODUCT PRODUCED WITH AI ASSISTANCE 15
1. Methodology *16*
2. Results of Human Simulations *19*
3. Comparing Human and AI Approaches to Simulations *22*
III. IMPLICATIONS AND QUESTIONS **25**
A. POTENTIAL IMPLICATIONS OF JUNOR LAWYERS' DIFFICULTY ASSESSING AI-GENERATED CONTENT 26
B. TOWARDS A BETTER UNDERSTANDING OF HOW AI USE IMPACTS LEGAL REASONING ... 29

Introduction

Robust evidence indicates that, as of mid-2025, generative AI has already attained the technological sophistication necessary to substantially transform legal practice. Empirical studies reveal that even outdated AI tools enable junior lawyers to complete various legal tasks 10% to 35% faster.[1] More recent AI tools preserve these efficiency benefits while also substantially enhancing the quality of junior lawyers' legal work.[2] Field studies echo these findings, indicating that attorneys report meaningful productivity gains after adopting AI.[3] And numerous benchmarking studies demonstrate that AI-generated responses to legal tasks frequently surpass the quality of the answers that are supplied by human lawyers who are unaided by AI.[4]

---

[1] Jonathan H. Choi, Amy Monahan, & Daniel Schwarcz, Lawyering in the Age of Artificial Intelligence, 109 Minn. L. Rev. 147, 150 (2024).

[2] Schwarcz, Daniel and Manning, Sam and Barry, Patrick James and Cleveland, David R. and Prescott, J.J. and Rich, Beverly, AI-Powered Lawyering: AI Reasoning Models, Retrieval Augmented Generation, and the Future of Legal Practice (March 02, 2025). Minnesota Legal Studies Research Paper No. 25-16, Available at SSRN: https://ssrn.com/abstract=5162111 or http://dx.doi.org/10.2139/ssrn.5162111.

[3] Miriam Kim & Colleen V. Chien, Generative AI and Legal Aid: Results from a Field Study and 100 Use Cases to Bridge the Access to Justice Gap, 57 Loyola LA Law Rev. 903, 904 (2025); Colleen V. Chien, Miriam Kim, Akhil Raj, & Rohit Rathish, How Generative AI Can Help Address the Access to Justice Gap Through the Courts, 57 Loyola LA Law Rev. 850 (2025).

[4] Vals Legal AI Report, (2025), https://www.vals.ai/vlair. Daniel Martin Katz, Michael

Despite these AI-enabled possibilities, evidence also shows that AI has yet to fundamentally alter legal markets or doctrines as of mid-2025. Although a growing percentage of lawyers report using AI in their practice,[5] the majority still largely resist AI.[6] Moreover, most attorneys who do use AI confine their use of this technology to narrow applications or limited contexts, like legal research. Broad industry indicators, such as demand and pricing for legal services, law school application rates, and lawyer compensation, have not significantly shifted yet in response to AI technology's rapid evolution in the last several years.[7]

To be sure, the gap between AI's technological capabilities and its actual impact on lawyering and judging likely reflects familiar delays in the diffusion of new technologies—delays that are neither unique to AI nor specific to the legal domain.[8] It also no doubt reflects a range of socio-legal and institutional factors, from regulatory constraints to concerns about how AI use interacts with process-based values long associated with the law.[9]

This Essay argues, however, that another important and underappreciated contributor to the gap between AI's current capabilities and its adoption is the difficulty that lawyers sometimes face in deeply understanding AI-generated output. This lack of deep understanding can undermine the ability of lawyers to evaluate the quality of AI-generated output and to perform related tasks, such as discussing issues with clients or judges, or integrating insights across discrete tasks. While these limitations may diminish over time, we suggest that they also may reflect fundamental aspects of how lawyers reason—particularly when they are relatively junior and inexperienced. If so, they are likely to shape how AI is used and adopted within the legal profession in the near to medium term.

James Bommarito, Shang Gao & Pablo Arredondo, GPT4 Passes the Bar Exam, PHIL. TRANS. R. SOC. A, Apr. 15, 2024, at 1, 3–5. cf. Eric Martínez, Re-Evaluating GPT-4's Bar Exam Performance, 1 Artificial Intelligence and Law 1 (2024).

[5] See, e.g., Bob Ambrogi, ABA Tech Survey Finds Growing Adoption of AI in Legal Practice, with Efficiency Gains as Primary Driver (March 7, 2025), https://www.lawnext.com/2025/03/aba-tech-survey-finds-growing-adoption-of-ai-in-legal-practice-with-efficiency-gains-as-primary-driver.html;

[6] See ABA's 2024 TechReport

[7] See Part I.B, infra.

[8] See, e.g., Hall, Bronwyn H., and Beethika Khan. "Adoption of new technology." (2003); Jennifer F. Reinganum, On the Diffusion of New Technology: A Game Theoretic Approach, *The Review of Economic Studies*, Volume 48, Issue 3, July 1981, Pages 395–405, https://doi.org/10.2307/2297153

[9] Tom Tyler, Why People Obey the Law; Metikoš, Ljubiša and van Domselaar, Iris, Procedural Justice and Judicial AI; Substantiating Explainability Rights with the Values of Contestation (May 05, 2025). Available at SSRN: https://ssrn.com/abstract=5242905 or http://dx.doi.org/10.2139/ssrn.5242905.

To advance these arguments, we discuss results from several recent experiments that are each described more fully, but with different emphases, in other articles.[10] The first such experiment is a randomized controlled trial showing that providing upper-level law students with access to leading AI tools improves both the speed at which they complete a range of realistic lawyering tasks and, in many cases, the quality of the resulting work. More importantly for present purposes, however, the results also indicate that participants generally failed to recognize when, or the extent to which, their use of AI improved the quality of their work. This failure, moreover, had nothing to do with the presence or absence of sources, a risk that can be easily safeguarded against by requiring lawyers to check AI-generated sources.[11]

A second set of experiments helps illuminate why lawyers may struggle to evaluate the quality of AI-generated output when addressing unfamiliar legal problems. In a series of simulations, we observed and analyzed the workflows and cognitive strategies of trained upper-level law students as they provided entity formation advice to small business clients.[12] While our primary aim was to gather detailed data on legal reasoning to inform the design of AI tools, the simulations revealed that students relied on highly dynamic, recursive, and individualized reasoning processes—even when they had helped create structured task prompts in advance. These findings suggest that, when confronted with unfamiliar legal issues, junior lawyers often depend on non-linear and exploratory reasoning to make sense of the task at hand. By contrast, the structured and linear workflows that AI generated in response to these tasks, and that are typical of AI output more generally, may bypass or inhibit these critical cognitive processes. Although these results are exploratory and may not extend to more experienced lawyers, they underscore a significant and largely underappreciated risk: that AI tools, when used for complex or unfamiliar legal tasks, may impair the ability of lawyers to develop, improve, explain, or extend their legal work product.

Notably, these tentative conclusions stand in sharp contrast to domains like coding, where even relatively inexperienced users can typically evaluate the quality of AI output using straightforward methods like compiling code.

---

[10] The primary results of each experiment are reported in other papers. Here, however, we focus on distinct features of these experiments that were not our primary focus in either of the other papers.

[11] Schwarcz et al., supra note 2.

[12] Debarati Das, Khanh Chi Le, Ritik Sachin Parkar, Karin De Langis, Brendan Madson, Chad M. Berryman, Robin M. Willis, Daniel H. Moses, Brett McDonnell, Daniel Schwarcz, Dongyeop Kang, LawFlow: Collecting and Simulating Lawyers' Thought Processes, https://arxiv.org/abs/2504.18942.

Our analysis is therefore distinct from—though related to—emerging evidence that reliance on AI can impede long-term learning by limiting opportunities for skill development among novice users.[13] Instead, we suggest that features intrinsic to legal reasoning can make it difficult for junior lawyers to accurately self-assess or fully comprehend assigned legal problems when they use AI to help resolve them.

In our final section, we outline future directions for testing our tentative conclusions, including potential experiments that more precisely evaluate how lawyers understand legal problems after working with—or without—the assistance of AI. We also consider how, if our assessments are correct, they may shape AI's future trajectory in legal practice. In particular, we suggest that firms may want to limit AI usage among junior attorneys to specific tasks that do not implicate complex legal reasoning, while encouraging AI usage among more expert lawyers on tasks for which they are well equipped to quickly and accurately evaluate the quality, limitations, and implications of AI-generated materials. We also suggest that if junior lawyers consistently struggle to deeply understand AI-generated outputs or assess their quality, this may prompt a broader shift in law firm business models.

I. The Legal AI Gap

A. AI's Current Power to Transform Legal Practice

A growing body of evidence demonstrates that generative AI already has the technological capability to significantly transform legal practice. Particularly strong support for this conclusion comes from two recent randomized controlled trials, in which upper-level law students were tasked with completing a diverse set of realistic lawyering tasks, either with or without the assistance of AI.

In the first study, conducted at the University of Minnesota Law School, students who were given access to GPT-4—a model widely regarded as a frontier system between March 2023 and early 2024, though now considered largely obsolete—completed four assigned legal tasks approximately 5% to 38% faster, on average, than students who did not use AI.[14] However, access to GPT-4 had no statistically significant effect on the quality of participants' work product.

---

[13] See Lee, Hao-Ping Hank, et al. "The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers." (2025).

[14] Jonathan H. Choi, Amy Monahan, & Daniel Schwarcz, Lawyering in the Age of Artificial Intelligence, 109 Minn. L. Rev. 147, 150 (2024).

Participants in the second study, conducted at the University of Minnesota and the University of Michigan Law Schools, completed six realistic lawyering tasks, with access to AI tools varying by task. For two tasks, they used the first publicly available AI reasoning model (o1 preview); for two others, they used a legal technology tool built on GPT-4 and enhanced with Retrieval-Augmented Generation (RAG); and for the final two tasks, no AI assistance was provided. Participants using both of the AI tools achieved comparable gains in speed as in the first study testing GPT-4.[15]

But unlike the first study, both AI tools also led to statistically significant improvements in work quality on four out of six assignments. Blind grading by professional legal writing experts, using predefined rubrics, indicated that AI-assisted submissions were 8% to 28% higher in quality on average, depending on the task and tool used. Although use of the reasoning model was associated with a higher incidence of hallucinated sources, use of the RAG-enhanced GPT-4 tool resulted in roughly the same number of hallucinated sources as in the control group. At the same time, participants with access to the reasoning model saw greater improvements in the analytical depth of their work than those using the RAG tool layered on top of GPT-4.

Two tables from this second study (reproduced below, and re-numbered) highlight its central findings: the first shows the effect of each AI model (o1-preview and Vincent AI) on the overall quality of the legal work product produced by study participants across the six assignments, and the second reports the relative changes in the speed with which participants completed these assignments. In both cases, the AI-assisted groups are compared to a control group that completed the same tasks without AI assistance.

[15] Schwarcz, Daniel and Manning, Sam and Barry, Patrick James and Cleveland, David R. and Prescott, J.J. and Rich, Beverly, AI-Powered Lawyering: AI Reasoning Models, Retrieval Augmented Generation, and the Future of Legal Practice (March 02, 2025). Minnesota Legal Studies Research Paper No. 25-16, Available at SSRN: https://ssrn.com/abstract=5162111 or http://dx.doi.org/10.2139/ssrn.5162111.

Table 1: Treatment Effects on Total Score Across Tasks

Task	Control Mean	Model	Effect	SE	% Change	N
Draft Client Email	19.341	Vincent	2.926*	(1.580)	+15.1%	136
		o1-preview	1.829	(1.435)	+9.5%	136
Draft Legal Memo	16.909	Vincent	2.277*	(1.212)	+13.5%	126
		o1-preview	3.988***	(1.194)	+23.6%	126
Analysis of Complaint	24.400	Vincent	1.941*	(1.098)	+8.0%	127
		o1-preview	2.484**	(1.207)	+10.2%	127
Draft NDA	26.395	Vincent	0.066	(0.824)	+0.3%	127
		o1-preview	−1.106	(0.862)	−4.2%	127
Draft Motion to Consolidate	17.489	Vincent	2.093*	(1.244)	+12.0%	127
		o1-preview	4.921***	(1.079)	+28.1%	127
Draft CNC Enforcement Letter	19.564	Vincent	−1.746	(1.697)	−8.9%	126
		o1-preview	4.087***	(1.560)	+20.9%	126

Notes: Effects shown as absolute increase relative to No AI control group. Percent changes calculated relative to control group mean. Robust standard errors in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Sample size (N) represents the number of observations used in the regression.

Table 2: Treatment Effects on Time Spent Across Tasks

Task	Control Mean	Model	Effect	SE	% Change	N
Draft Client Email	50.302	Vincent	−7.147***	(2.544)	−14.2%	135
		o1-preview	−6.111***	(2.277)	−12.1%	135
Draft Legal Memo	183.909	Vincent	−30.374***	(11.302)	−16.5%	127
		o1-preview	−25.984**	(12.097)	−14.1%	127
Analysis of Complaint	107.103	Vincent	−39.512***	(5.226)	−36.9%	126
		o1-preview	−29.917***	(5.715)	−27.9%	126
Draft NDA	96.302	Vincent	−5.052	(10.443)	−5.2%	128
		o1-preview	−13.302	(9.296)	−13.8%	128
Draft Motion to Consolidate	95.800	Vincent	−17.405**	(8.135)	−18.2%	128
		o1-preview	−17.500**	(8.476)	−18.3%	128
Draft CNC Enforcement Letter	110.950	Vincent	−38.677***	(8.128)	−34.9%	127
		o1-preview	−29.136***	(7.710)	−26.3%	127

Notes: Effects shown as absolute increase relative to No AI control group. Percent changes calculated relative to control group mean. Robust standard errors in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Sample size (N) represents the number of observations used in the regression.

These randomized controlled trials are particularly valuable for assessing AI's current potential impact on legal practice because they directly measure how providing AI tools affects human performance.[16] Given that licensing regulations, ethical norms, and established professional practices currently require licensed human lawyers to affirmatively verify and scrutinize AI-generated output, assessing AI's impact on human lawyers represents the

---

[16] Fabrizio Dell'Acqua et al., Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality (Sept. 15, 2023) (unpublished manuscript) (on file with authors) [hereinafter *Navigating the Jagged Technological Frontier*].

most relevant measure of AI's present capacity to influence legal practice.[17] Moreover, employing expert human graders who blindly assessed output using detailed rubrics enhances the reliability and validity of the findings regarding AI's impact on work quality.

At the same time, the primary limitation of these studies is their reliance on upper-level law students rather than practicing attorneys as study subjects. On one hand, there is good reason to suspect that upper-level law students from top law schools are similar to junior attorneys with limited experience in both their ability to use AI and to complete basic realistic lawyering tasks of the type tested in these experiments. On the other hand, the study participants are quite likely different from experienced lawyers with a deep expertise on the tasks that they perform regularly for clients.

Other studies employing different methodologies consistently also point to AI's current technological capacity to transform lawyering. Benchmarking studies, for instance, frequently find that AI-generated output matches or exceeds the quality of work produced by human lawyers.[18] One particularly prominent recent example is a benchmarking analysis conducted by the independent firm Vals.AI, which evaluated four legal AI tools—Harvey Assistant, CoCounsel (Thomson Reuters), Vincent AI (vLex), and Oliver (Vecflow)—across seven realistic legal tasks.[19] The AI outputs were compared against a "Lawyer Baseline" created by an "average," "appropriately skilled" group of attorneys who were tasked with performing identical legal assignments. According to an automated "LLM-as-judge" scoring system, at least one AI tool surpassed average human lawyer performance on most tasks. Other benchmarking studies echo these results, finding that AI-generated outputs frequently outperform human test-takers on law school and bar exams.[20]

---

[17] Even apart from these professional guidelines, empirical evidence highlights a persistent client preference for documents believed to be human-generated, consistent with the well-documented "algorithmic aversion" phenomenon. Harasta et al. (2024) – *"It Cannot Be Right If It Was Written by AI"*. See also Aileen Nielsen, Stavroula Skylaki, Milda Norkute, & Alexander Stremitzer, *Building A Better Lawyer: Experimental Evidence That Artificial Intelligence Can Increase Legal Work Efficiency*, 21 J. EMP. LEGAL STUDS. 979 (2024).

[18] See generally Zhiwei Fei, Xiaoyu Shen, Dawei Zhu, Fengzhe Zhou, Zhuo Han, Alan Huang, Songyang Zhang, Kai Chen, Zhixin Yin, Zongwen Shen, Jidong Ge, and Vincent Ng. 2024. LawBench: Benchmarking Legal Knowledge of Large Language Models. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, pages 7933–7962, Miami, Florida, USA. Association for Computational Linguistics.

[19] Vals Legal AI Report, (2025), https://www.vals.ai/vlair.

[20] Daniel Martin Katz, Michael James Bommarito, Shang Gao & Pablo Arredondo, GPT4 Passes the Bar Exam, Phil. Trans. R. Soc. A, Apr. 15, 2024, at 1, 3–5. cf. Eric Martínez, Re-Evaluating GPT-4's Bar Exam Performance, 1 Artificial Intelligence and Law

Of course, these benchmarking studies possess their own limitations. Chief among them is their comparison of AI-generated work to human-generated work, which fails to provide direct evidence regarding the key question: the effect of providing human lawyers with AI-based tools.[21] While one might reasonably infer that AIs capable of producing high-quality work would substantially improve lawyers' productivity, this is not self-evident. In some contexts, for instance, verifying AI-generated work may require more effort than independently producing it from scratch. A second significant limitation of benchmarking studies involves their methodologies for evaluating output quality. Benchmarking studies like the one conducted by Vals.AI often rely on AI systems to grade AI and human output, even though there is no evidence of which we are aware suggesting that AI-assigned grades accurately and consistently reflect the quality of legal work product.

A third methodological approach, which also points to the potential transformative impact of AI on lawyering, involves field studies that directly evaluate the impact of providing AI tools to practicing attorneys. For example, one study provided GPT-4 access to approximately 100 legal aid professionals, finding that 90% reported increased productivity and 75% intended to continue using AI.[22] The primary advantage of this approach lies in its direct relevance to actual legal practice and real-world user populations—practicing attorneys rather than law students. However, the principal drawback of field studies is their reliance on subjective, self-reported measures rather than objective performance metrics.

Though each type of study discussed above—randomized controlled trials, benchmarking analyses, and field experiments—has distinct methodological limitations, collectively they provide strong and consistent evidence that generative AI already possesses the technological capability to dramatically reshape legal markets. And, indeed, surveys of lawyers and law firms consistently show that most lawyers do indeed expect AI to substantially transform the practice of law in the near future.[23]

---

1 (2024); Jonathan H. Choi & Daniel Schwarcz, AI Assistance in Legal Analysis: An Empirical Study, 73 J. Legal Ed. 384 (2025); Jonathan H. Choi, Amy B. Monahan, & Kristin Hickman, & Daniel Schwarcz, ChatGPT Goes to Law School, 71 J. Legal Ed. 387 (2022).

[21] See, e.g., Nicole Yamane, Artificial Intelligence in the Legal Field and the Indispensable Human Element Legal Ethics Demands, 33 GEO. J. LEGAL ETHICS 877, 882 (2020); pages 7-8 of Frazier and Rozenshetin on Large Language Scholarship

[22] Miriam Kim & Colleen V. Chien, Generative AI and Legal Aid: Results from a Field Study and 100 Use Cases to Bridge the Access to Justice Gap, 57 Loyola LA Law Rev. 903, 904 (2025); Colleen V. Chien, Miriam Kim, Akhil Raj, & Rohit Rathish, How Generative AI Can Help Address the Access to Justice Gap Through the Courts, 57 Loyola LA Law Rev. 850 (2025).

[23] See, e.g., WoltersKluwer, Legal innovation: Seizing the future or falling behind?

B. Limited Evidence of AI-Driven Legal Market Transformations

Recent evidence makes clear that, despite its transformative potential, AI is currently an ancillary rather than central feature of legal practice as of mid-2025 for most lawyers and law firms. Multiple sources indicate that while the percentage of lawyers using AI is indeed accelerating quickly,[24] a majority of lawyers still do not regularly use generative AI in their legal work.[25] And at the organizational level, roughly 78 percent of law firms reported in one 2025 survey that they do not deploy any AI tools across the firm.[26]

Among the minority of lawyers who do use AI in their work, usage is often tightly circumscribed. A 2024 ABA report describes adoption as "still relatively limited to certain functions," with the most common tasks being routine correspondence (54 percent), brainstorming (47 percent) and general research (46 percent).[27] Far fewer surveyed attorneys rely on the technology to assist with complex advisory work, negotiation strategy or court argument.[28]

Broader legal market metrics likewise show little disruption from AI. For instance, AI has not yet significantly impacted law firms' billing models,[29]

(2024); CLIO, Legal Trends Report (2024); Robert J. Couture, The Impact of Artificial Intelligence on Law Firms' Business Models, Harvard Law School Center on the Legal Profession

[24] https://www.lawnext.com/2025/04/an-ai-assisted-look-at-four-new-surveys-on-ai-adoption-in-law-how-do-they-compare-differ.html

[25] See Robert Abrbogi, ABA Tech Survey Finds Growing Adoption of AI in Legal Practice, with Efficiency Gains as Primary Driver, https://www.lawnext.com/2025/03/aba-tech-survey-finds-growing-adoption-of-ai-in-legal-practice-with-efficiency-gains-as-primary-driver.html#:~:text=There%20has%20been%20a%20significant,Association%E2%80%99s%20Legal%20Technology%20Survey%20Report; 13.5% of Inhousers: AI Tools 'Essential' – Survey 30th April 2025 artificiallawyer Inhouse AI Use Comments Off ("a new survey by Spotdraft has found that 13.5% of inhouse respondents said that they use AI tools daily and that this approach is essential for their work. Plus, an additional 28.2% said they used AI tools 'a few times per week."; AffiniPay Launches 2025 Legal Industry Report: Embracing Technology, Financial Wellness, and the Future of Legal Work

[26] https://www.embroker.com/blog/risks-and-benefits-of-ai-for-lawyers/#:~:text=However%2C%20despite%20the%20increasing%20opportunities,unintended%20consequences%2C%20and%20security%20vulnerabilities

[27] Ambrogi, supra note 25.

[28] Id.

[29] CLIO Report (noting a 6% increase in the number of firms adopting flat fee billing in the past year, but recognizing that this trend largely pre-dates the rise of AI and is only partly attributable to AI adoption); https://www.bestlawfirms.com/articles/billable-hours-endure-law-firms-expand-offerings/6208?utm (

billing rates,[30] size,[31] or overall profitability.[32] Surveys find no evidence of AI-driven job losses inside corporate legal departments or law firms, and overall lawyer employment remains robust. In short, the labor market for lawyers looks much the same today as it did before ChatGPT's debut. Reflecting these economic realities, legal education pipelines remain strong: law-school application volume for the 2025 academic year actually rose by 21 percent in the U.S., marking that year as one of the most competitive ever for law school admissions.[33]

Generative AI has likewise left only faint fingerprints on judicial decision-making. To be sure, judges and court systems have used generative AI tools for narrow purposes such as facilitating legal research.[34] However, remarkably few judges in the U.S. have publicly admitted to using generative AI to help draft opinions, understand the parties' briefing, or digest other relevant case-specific documents.[35] The most notable caveat is that one federal appellate judge in the U.S. suggested, in a pair of exploratory concurring opinions, that generative AI tools could helpfully supplement traditional interpretive tools in the context of an insurance coverage dispute and a dispute about the application of the federal sentencing guidelines.[36]

---

[30] Billing Rates Keep Increasing. GCs Question Whether ... (2/25/25) https://www.law.com/americanlawyer/2025/02/24/billing-rates-keep-increasing-gcs-question-whether-their-growth-is-sustainable/?slreturn=2025050593033

[31]

[32]

[33] Sara Randazzo, The Competition to Get Into Law School Is Brutal This Year: Applications are surging as students seek stability in a difficult job market, WSJ, March 15, 2025 5:30 am ET. See also Lisa Larrimore Ouellette, Amy Motomura, Jason Reinecke, and Jonathan S. Masur. 2025. Can AI Hold Office Hours?

[34] See Ohio Northern University Law Review, Vol. 50 [2024], Iss. 3, Art. 2; Marcus W. Reinkensmeyer & Raymond L. Billotte, Artificial Intelligence (AI): Early Court Project Implementations and Emerging Issues, The National Association For Court Management (Aug. 2019), https://thecourtmanager.org/articles/artificial-intelligence-ai-early-courtproject-implementations-and-emerging-issues/.

[35] Gary E. Marchant, *AI in Robes: Courts, Judges, and Artificial Intelligence,* 50 OHIO N U L REV. 473 (2024); See JIC ADVISORY OPINION 2023-22 at 5 (in drafting judicial opinions, "a judge may use AI for research purposes but may not use it to decide the outcome of a case. The use of AI in drafting opinions or orders should be done with extreme caution). See also Eric A. Posner and Shivam Saran. 2025. Judge AI: Assessing Large Language Models in Judicial Decision Making.

[36] Snell v. United Specialty Insurance Co., 102 F. 4th 1208 (11th Cir. 2024) (Newsom, J., concurring) (using ChatGPT to explore the appropriate interpretation of an insurance policy's use of the word "landscaping" in the context of the installation of a trampoline); United States v. Deleon, 116 F. 4th 1260, 1270 (11th Cir. 2024) (Newsom, J., concurring) (using ChatGPT to consider the meaning of the phrase "physically restrained" in the U.S. Sentencing Guidelines). See generally Waldon, Brandon and Schneider, Nathan and Wilcox, Ethan and Zeldes, Amir and Tobia, Kevin, Large Language Models for Legal Interpretation?

Additionally, judges in several countries—including India, Dubai, Peru, Mexico, and the United Kingdom—have publicly acknowledged using generative AI tools like ChatGPT to assist in drafting judicial opinions or orders.[37]

In sum, while there continues to be significant speculation that AI will dramatically reshape the legal landscape soon, these predictions have not yet come to pass for most lawyers. While AI is increasingly one tool that many lawyers use for certain specific tasks, broad adoption of AI across a wide-ranging swath of legal tasks and workflows remains uncommon.

II. The Difficulty Human Lawyers Face Evaluating the Quality of AI Output

Even transformational technologies are rarely adopted broadly or immediately. Adoption delays commonly stem from factors such as cost, lack of awareness, institutional or regulatory inertia, steep learning curves, and network effects. These familiar barriers are undoubtedly slowing uptake of AI among lawyers and judges.

This Part, however, argues that an additional factor—distinctive to legal practice and unlikely to dissipate quickly—also helps explain the profession's hesitancy towards broadly adopting AI. In particular, we argue that inexperienced lawyers can struggle to evaluate the quality and reliability of AI-generated legal output in ways that preserve AI's potential efficiency gains. Crucially, this challenge is not primarily about hallucinated content, which is increasingly straightforward to minimize. Rather, we suggest that this difficulty reflects deeper features of legal reasoning in unfamiliar terrain—namely, the link between deep legal reasoning and recursive, individualized, and often messy analytical processes. Although these results are exploratory and may not generalize to more experienced lawyers, they suggest that using AI tools can diminish lawyers' ability to critically evaluate their own work product or to build upon it when tackling related tasks.

A. The Difficulty Junior Lawyers Face Evaluating the Quality of AI Output

---

Don't Take Their Word for It (February 03, 2025). Georgetown Law Journal, Vol. 114 (forthcoming), Available at SSRN: https://ssrn.com/abstract=5123124 or http://dx.doi.org/10.2139/ssrn.5123124

[37] Marchant, supra note 35, at 486; see also John Zhuang Liu and Xueyao Li. 2024. How Do Judges Use Large Language Models? Evidence from Shenzhen. Journal of Legal Analysis, 16(1):235–262.

Since the public release of ChatGPT in late 2022, many lawyers and judges have expressed concern that AI-assisted legal work may be comparatively low quality. These concerns have centered primarily on the risk that generative AI systems can "hallucinate" legal authorities, generating plausible but fictitious cases, statutes, or other sources.[38] Several high-profile incidents, in which court filings cited entirely invented legal authorities, intensified these fears.[39] In response, some judges began requiring disclosure of AI use, bar associations emphasized lawyers' ethical duty to verify all cited materials, and many lawyers and clients rejected AI tools outright due to their potential to produce inaccurate or fabricated content.[40]

In recent months, concerns about AI hallucinating legal authorities have receded. This shift is partly due to the recognition that lawyers have always made errors in citing cases and in characterizing their holdings. AI-generated hallucinations are in that way not categorically different from traditional lawyering mistakes. More significantly, both legal-specific and general-purpose AI tools now provide increasingly reliable methods for grounding their output in authentic legal sources, and enabling lawyers to verify the accuracy of cited materials directly. Indeed, the randomized controlled trial discussed in Part I demonstrated that RAG-enabled AI tools could largely eliminate hallucinated sources in this way.

And yet, new evidence suggests that the risk of low-quality AI-supported legal work cannot be dismissed merely by requiring lawyers to check the veracity of AI-generated citations. To the contrary evidence from several experiments, including the randomized controlled trial discussed in Part I, suggests quite the opposite.

Not surprisingly, the data from the randomized controlled trial testing the

---

[38] *See* 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 legal analysis of large language models); Varun Magesh, Faiz Surani, Matthew Dahl, Mirac Suzgun, Christopher D. Manning, Daniel E. Ho, *Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools* (May 30, 2024), at https://arxiv.org/abs/2405.20362

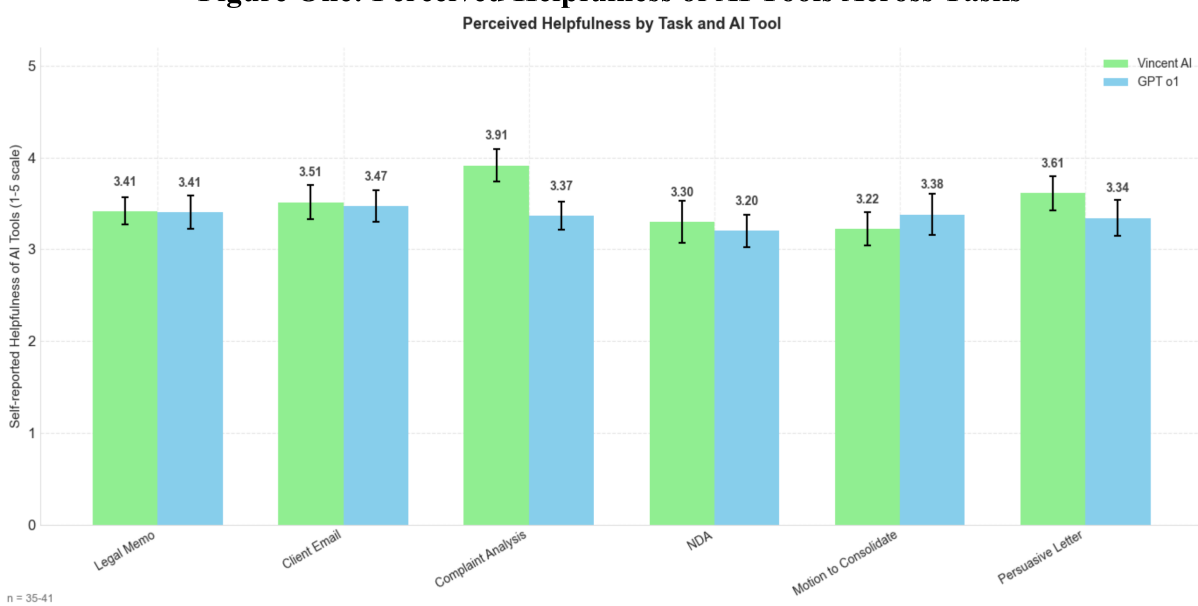
[39] Benjamin Weiser, *Here's What Happens when Your Lawyer Uses ChatGPT*, N.Y. TIMES (May 27, 2023), https://www.nytimes.com/2023/05/27/ nyregion/avianca-airline-lawsuit-chatgpt.html; Sara Merken, *New York Lawyers Sanctioned For Using Fake ChatGPT Cases In Legal Brief*, REUTERS, June 26, 2023; Larry Neumeister, *Lawyers Submitted Bogus Case Law Created By ChatGPT. A Judge Fined Them $5,000*, AP NEWS, June 22, 2023.

[40] ST. BAR CAL. STANDING COMM. ON PRO. RESP. AND CONDUCT, PRACTICAL GUIDANCE FOR THE USE OF GENERATIVE A.I. IN THE PRACTICE OF LAW 3 (2023), https://www.calbar.ca.gov/Portals/0/documents/ethics/Generative-AI-Practical Guidance.pdf; Jon Garon, *"Ethics 3.0—Attorney Responsibility in the Age of Generative AI,"* The Business Lawyer, Am. Bar Assoc (2024).

impact of providing law students with reasoning and RAG AI models (contained in Part I) show that the usefulness of the two tested AI tools varied significantly across tasks. Most notably, as evident in Table One, neither of the tested AI tools improved the quality of participants' work product on one of the six tested assignments: an assignment requiring the drafting of a Non-Disclosure Agreement (NDA).[41] Similarly, as reflected in Table Two, both AI tools significantly increased speed on five of the six tasks, with the NDA assignment again standing out as the sole task where AI access produced no measurable improvement in speed.[42] The results also suggest that the two tested AI tools were especially effective for several assignments. In particular, Table Two shows that the AI reasoning model (o1-preview) yielded especially sizable and statistically significant improvements in work quality for three assignments—the legal memo, the motion to consolidate, and the persuasive letter.[43]

Now compare these results to those reflected in the post-experiment survey of participants regarding the helpfulness of the various AI tools across different tasks, which are reproduced below in Figure One.

Figure One: Perceived Helpfulness of AI Tools Across Tasks



The most striking aspect of the post-experimental survey responses is how little they align with the objective performance data. Participants

---

[41] Table 3
[42] Table 9
[43] Table 3

generally rated both AI tools as equally helpful across all six tasks, including the NDA Assignment, despite clear evidence that AI had no measurable impact on either the quality or speed of work on that task. Similarly, participants failed to recognize the extent to which o1-preview significantly improved the quality of their work on the three assignments for which it was most helpful—the legal memo, the motion to consolidate, and the persuasive letter. Taken together, these results suggest that participants were largely unaware of when and how AI meaningfully enhanced the quality of their performance across the six tasks. In some cases, they thought AI helped more than it did, and in other cases they failed to identify how much AI improved their submitted work product.

Importantly, these disconnects cannot be explained by hallucinated source materials that participants may have overlooked. In fact, graders found no evidence of hallucinated sources in any of the NDA assignments, which was the one assignment on which the availability of AI produced no measurable quality or speed gains. Meanwhile the highest incidence of hallucinated materials occurred in responses to the motion to consolidate—one of the tasks for which AI assistance was most beneficial.

A separate study provides mild additional support for the conclusion that junior lawyers may struggle to accurately assess the quality of AI-generated legal output. In this study, upper-level law students took both a real law school exam (without AI) and a different, second exam covering the same subject matter, but this time with access to GPT-4. Notably, students who performed best on the original, no-AI exam tended to fare significantly worse—by as much as twenty percentile points—when using AI. [44] While this result is consistent with multiple interpretations, it suggests that high-performing students often failed to recognize the limitations of AI-generated answers. If they had done so, their demonstrated competence on taking exams without AI should have enabled them to identify and correct deficiencies in the AI-assisted responses.

B. Explaining Why Human Lawyers May Have Difficulty Evaluating the Quality of Legal Work Product Produced with AI Assistance

To better understand why junior lawyers may struggle to evaluate the quality of AI-assisted legal work product, Das et al.[45] conducted a series of simulations in which trained upper-level law students performed a range of

---

[44] Choi & Schwarcz, AI Assistance, supra note 20.

[45] Das, D., Le, K.C., Parkar, R.S., De Langis, K., Madson, B., Berryman, C.M., Willis, R.M., Moses, D.H., McDonnell, B., Schwarcz, D. and Kang, D., 2025. LawFlow: Collecting and Simulating Lawyers' Thought Processes. arXiv preprint arXiv:2504.18942.

legal tasks related to business entity formation, with a particular focus on drafting agreements and advising small businesses. Subsection (1) outlines the key elements of their methodology, while Subsection (2) discusses the results. Subsection (3) then compares these results to AI-generated outputs, highlighting differences between how humans and AI systems approached the legal simulations.

1. Methodology

Das et al. began by recruiting four high-performing third-year law students from the University of Minnesota with relevant experience in business law. One of the students was serving as the student director of the law school's small business law clinic and assumed a similar leadership role in the simulated exercises.

Working with two law professors, the four students developed a preliminary "Human Task Plan" outlining the key steps they anticipated that lawyers would typically follow when advising a small startup on entity formation. These steps included tasks such as gathering information, making recommendations, drafting client memos, and evaluating the tax implications of various potential entity choices. Each task was further broken down into concrete, practice-aligned actions such as client communication, legal research, and document drafting.

Using this initial plan as a guide, the four students conducted ten pilot simulations, advising a variety of fictitious small businesses on entity formation. These scenarios were loosely modeled on real cases from the law school's small business clinic and involved simulated clients drawn from the group of law students, computer science collaborators, and faculty co-authors. Based on insights from these simulations, the law students and professors refined their framework into a final Human Task Plan, presented in Figure Two.

Figure Two: Human Task Plan

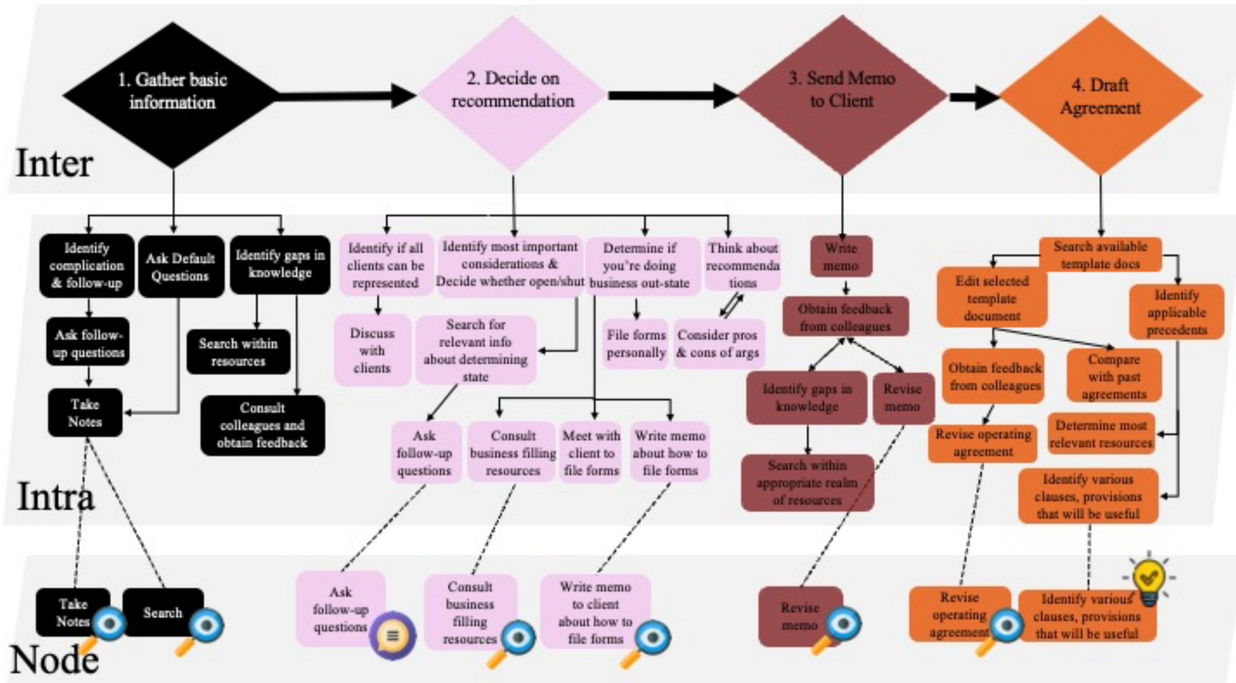
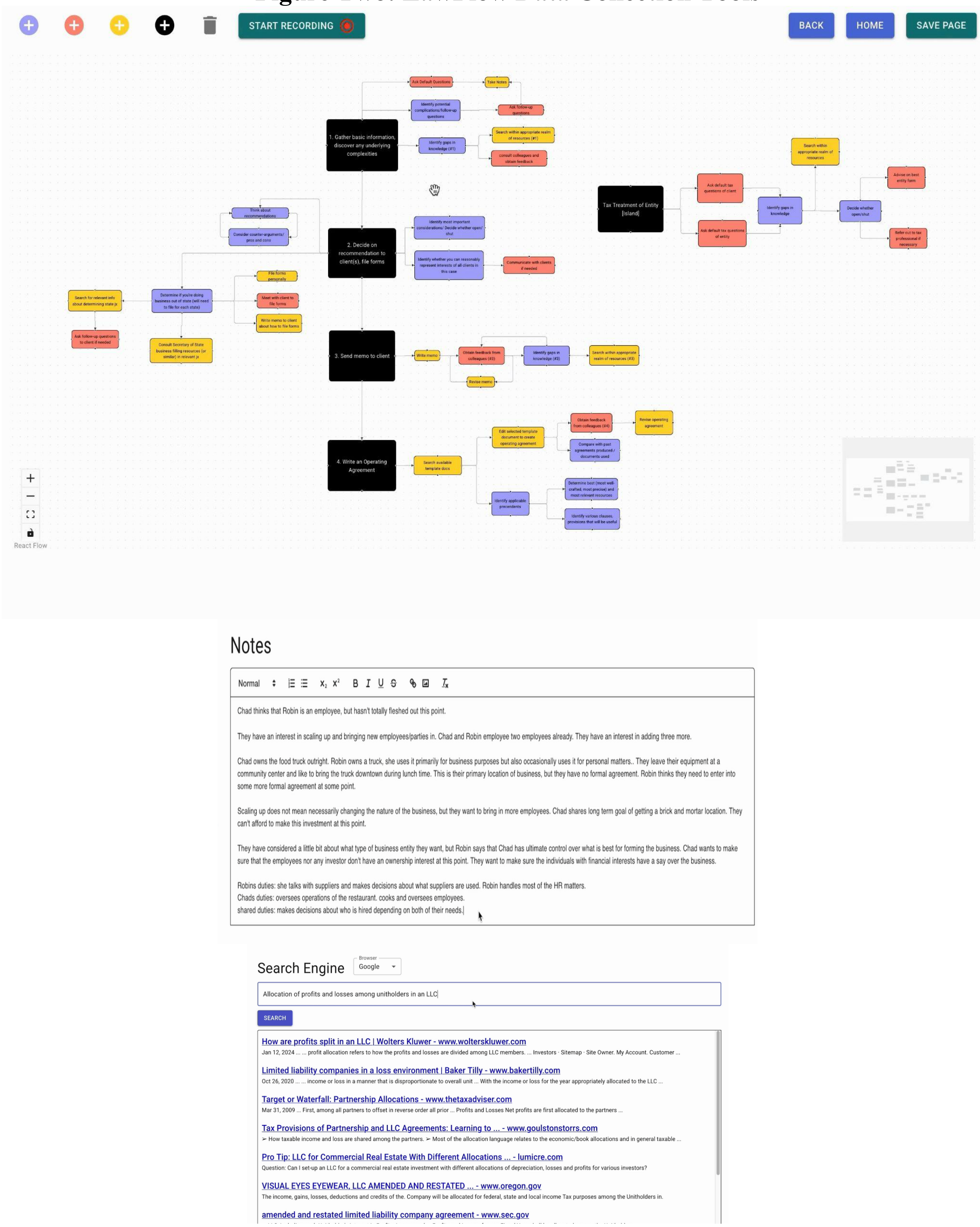


Figure 2: The human task plan illustrates the multi-level structure of the legal workflow behind drafting an Operating Agreement. The inter-subtask level (top row) outlines the primary stages of the process, from initial client intake to final document preparation. The intra-subtask level (middle row) decomposes each stage into finer-grained tasks and shows their interconnections. The node-level (bottom row) represents individual actions, annotated by cognitive modality: introspective (internal legal reasoning), interactive (client or colleague communication), and observable (use of external tools and resources). Thus, the task diagram captures the complexity, adaptivity, and tool-mediated nature of real-world legal reasoning.

While the co-authors from the law school refined the simulation process and finalized the Human Task Plan, the computer science co-authors developed a suite of specialized data collection tools to comprehensively record each element of the simulations. These tools integrated all resources participants would need during the exercises, including a library of standard legal forms, note-taking and legal research tools, audio recording capabilities, and document editing functions. Crucially, the system was designed to track how participants navigated among tasks and sub-tasks throughout the simulations, providing detailed insights into their workflow and decision-making processes. During the preliminary simulations, the law students became familiar with these tools and offered feedback to the computer science team to improve their usability and functionality. Figure 2 below provides screen shots of some of these data collection tools in their final form.

Figure Two: LawFlow Data Collection Tools



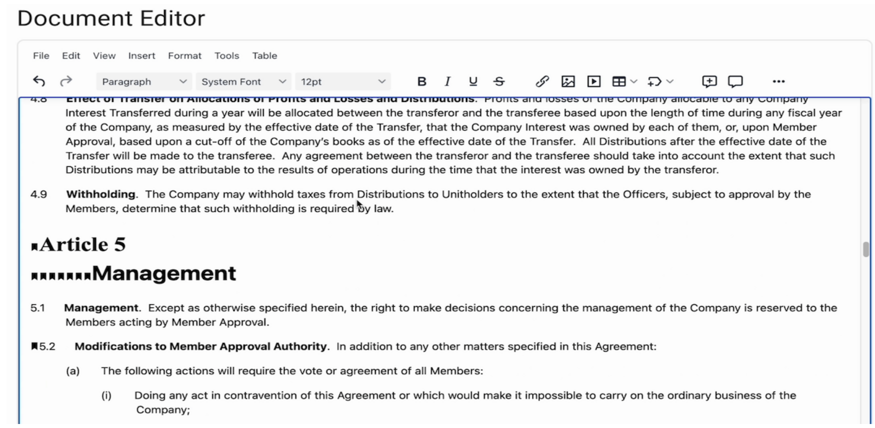


Figure 2: Views of the LawFlow data collection tool include 1) task diagram annotation, 2) note taking, 3) web search, and 4) document editing.

After completing the preliminary development process, the four law students conducted ten formal simulations that serve as the basis for the data reported below. Like the earlier pilot simulations, these exercises were loosely inspired—albeit to varying degrees—by actual cases previously handled by the law school's small business clinic.

2. *Results of Human Simulations*

Using the data collection tools and final Human Task Plan, Das et al. captured how the students navigated the ten simulated legal matters in real-time. They found that the students sharply deviated from the anticipated Human Task Plan, instead employing an adaptive, back-and-forth process rather than following a linear, step-by-step sequence. For instance, the students frequently paused to recheck facts, revisit earlier steps, and pose follow-up questions, particularly when client needs were ambiguous or information was incomplete. Two of these processes are illustrated in Figures 4 and 5 below.

Figure Four: Human Execution Workflow for Scenario One

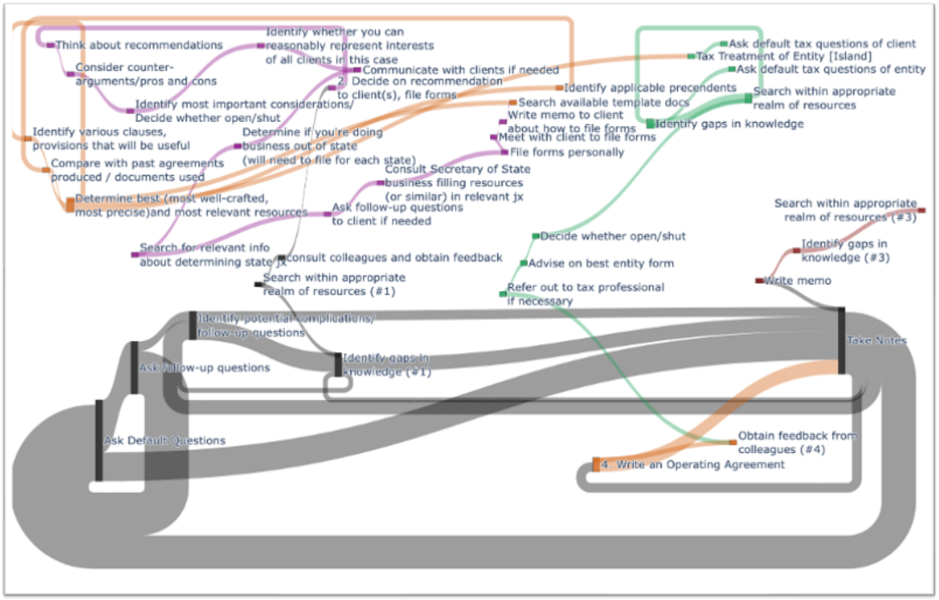


Figure 4: **Human execution workflow for Scenario 1** - A and B are avid anglers aiming to promote fishing in our area by hosting free educational workshops for children and adults. They plan to form a business to manage gear purchases, raffles, and possible future funding, but want to avoid complex formalities and do not intend to make a profit. They are interested in understanding what kind of startup agreement they need.

Figure Five: Human Workflow for Scenario Two

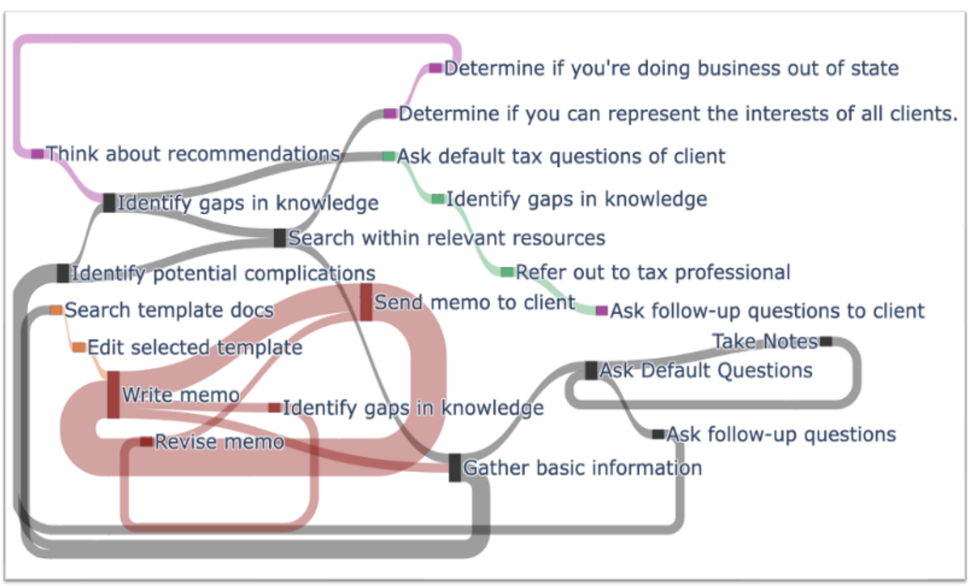


Figure 4: **Human execution workflow for Scenario 2** - A and B have created the chemical

composition of an organic, temperature-regulating soil. A and B need additional funding to produce an amount large enough to be tested on a greater scale & form a viable business. A and B's friend C has an eye for promising business ideas. She knows the two have a strong work ethic and believes that the soil will succeed if they can consistently produce enough to meet early testers' demands. C is willing to invest $2M for a 30% ownership interest, which A and B agree to. If all goes well, the three want to sell the soil's chemical composition to a biotech company.

In one scenario (Figure 4), where the client was uncertain about next steps, the workflow became more exploratory and recursive, with the student circling back to earlier tasks to clarify gaps. In contrast, the workflow was more streamlined in a more straightforward scenario (Figure 5), focusing on next steps and formalizing recommendations. Figure 6 provides a visualization of how tasks were executed across all ten of the simulations the humans conducted.

Figure Six: Overview of Human Workflows Across All Simulations

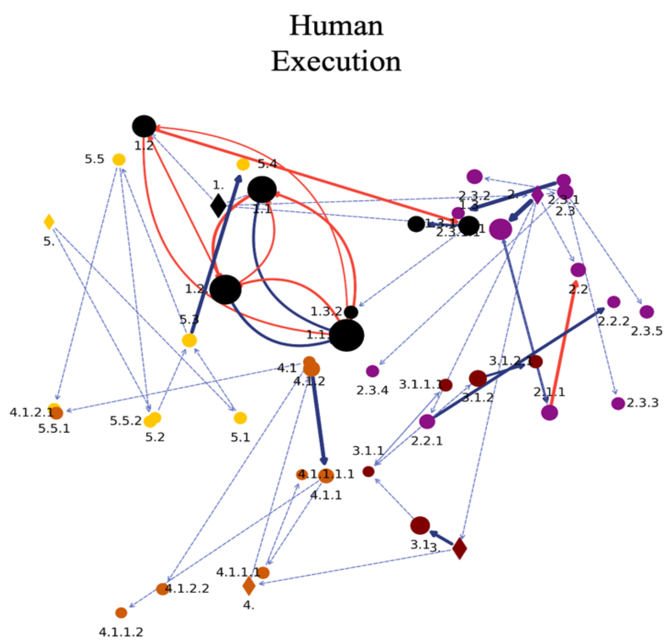


Figure 6: Finite State Machines (FSMs) illustrate human workflows across all of the simulations, representing as interconnected nodes (steps) and edges (transitions between steps) the aggregate data. In this diagram, *node size* reflects the frequency of revisits, and *edge thickness* indicates the likelihood of each transition. *Blue edges* represent planned transitions, while *red edges* show deviations where steps were executed outside the original task plan. *Solid edges* indicate executed transitions, and *dashed edges* denote planned but unexecuted steps.

Several key points are particularly evident from the visualization in

Figure Six. First, the relatively large node sizes for certain tasks reflect the tendency of the law students to continuously revisit key steps during their workflows. These revisits often were not anticipated based on the final human task plan they had helped develop earlier, as indicated by the red lines connecting the large nodes. At the same time, the law students often skipped planned steps, as reflected in the dashed lines in Figure Six. In sum, Figure Six shows selective, depth-oriented exploration of the anticipated task plan, with frequent targeted and often unplanned revisits to key tasks.

3. Comparing Human and AI Approaches to Simulations

To better understand how human navigation of these simulations compared to AI output, Das et al. used several reasoning AI models, including GPT-o1 and Deepseek-R1, to develop (i) an AI Task Plan, and (ii) an AI Workflow for each of the ten simulations on which data was reported above.

Mirroring the approach used in the human simulations, the researchers first developed a task plan by prompting LLMs to generate a detailed, step-by-step plan for handling business entity formation across a range of scenarios. To do so, the researchers prompted the models to consider the same ten pilot simulations used in the preliminary human simulations and to identify core legal tasks—such as client consultations, entity selection, agreement drafting, and tax planning. Like their human counterparts, the AI systems were directed to organize these tasks hierarchically into main tasks, sub-tasks, and further sub-steps as needed, producing a structured and comprehensive workflow.

The resulting AI Task Plan diverged in several important respects from the Human Task Plan shown in Figure Two. While the Human Task Plan selectively emphasized areas of uncertainty or high importance, the AI-generated plan took the form of an exhaustive, uniformly structured list of tasks. It treated each step as equally significant, lacking contextual cues to indicate which elements warranted deeper analysis or greater attention. Unlike the human-generated plan—which was designed to adapt to specific client needs and highlight critical decision points—the AI plan prioritized completeness over nuance. This contrast is illustrated in Figure 7. Human task plans begin with a small set of high-level goals that are incrementally broken down into more specific actions, supporting flexible navigation and responsiveness to client input. By contrast, the LLM-generated plans offer a flat enumeration of detailed tasks without a clear hierarchy, potentially making it harder for junior lawyers to identify which steps are most important or relevant.

Figure Seven: Human and AI Task Plans

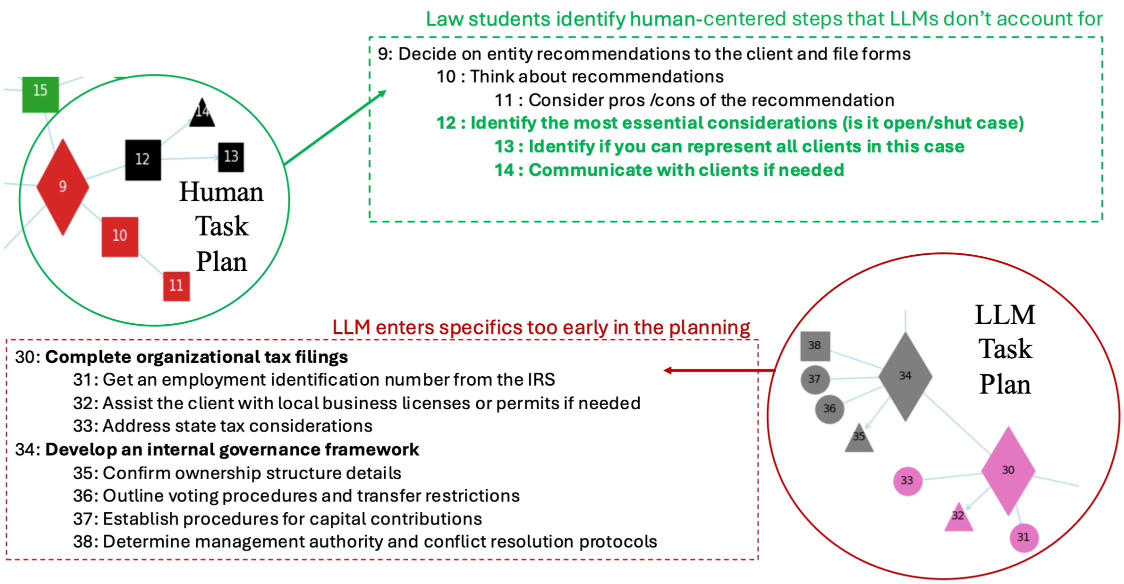


Figure 7: This figure depicts that law students plan by starting with a small set of high-level goals and anticipate that they may need to decompose them by identifying knowledge gaps or consulting with clients. By contrast, AI's task plan lays out a flat, exhaustive task list with little hierarchical structure and more focus on the nitty-gritty details.

As with the human simulations, Das et al. also attempted to generate AI workflows for the ten simulations on which human data was collected. To do so, they prompted the AI systems to act as a lawyer working through each of the ten simulated scenarios by interacting with clients or reflecting on the task internally. Each step was framed as a distinct decision point based on the Human Task Plan, allowing the AI to simulate real-time decision-making as new information emerged. The AI was asked to assess the relevance of each step, describe the lawyer's actions or dialogue, and suggest the next step in the process. This structure was intended to create a "chain-of-decision" reasoning structure, connecting discrete steps into a coherent workflow where each decision influenced the next. Unlike the AI Task Plan, this structure was designed to mimic how human lawyers might adapt their reasoning, potentially revisiting earlier steps or branching into related tasks based on evolving information. Figure 8 visualizes the aggregate results of the AI's workflows across the ten simulations.

Figure Eight: Overview of AI Workflows Across All Simulations

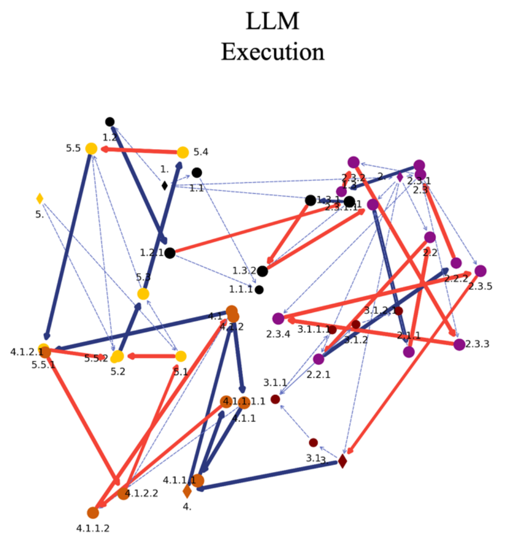


Figure 8: Finite State Machines (FSMs) illustrate AI workflows across all of the simulations, representing as interconnected nodes (steps) and edges (transitions between steps) the aggregate data. In this diagram, *node size* reflects the frequency of revisits, and *edge thickness* indicates the likelihood of each transition. *Blue edges* represent planned transitions, while *red edges* show deviations where steps were executed outside the original task plan. *Solid edges* indicate executed transitions, and *dashed edges* denote planned but unexecuted steps.

Comparing the human workflow in Figure 6 and the AI workflow in Figure 8 reveals several key differences. Relative to the human workflow, the AI's workflow remained mostly linear, methodically progressing from one task to the next without performing the adaptive, back-and-forth reasoning observed in the human workflows. Figure 8 makes this clear: the nodes are small, dashed edges are sparse, and sibling nodes are densely cross-linked—evidence that the AI is exhaustively pursuing many actions in parallel. This linearity contrasts sharply with human workflows, which display frequent, targeted revisits to key nodes and selective execution of parallel actions, illustrating how students adjusted their reasoning in response to evolving client needs or emerging information.

Of course, these differences between the human and AI task plans and execution graphs in these simulations are based only on the efforts of a small handful of law students tackling a select number of scenarios. Even so, when considered in light of the data discussed in Part A, they lead to several tentative conclusions about why junior lawyers like those in our simulations may struggle to evaluate the quality of AI-generated legal outputs. Specifically, these data suggest that the structural rigidity of AI task plans and workflows contrasts sharply with the reasoning processes of less experienced lawyers. When confronting unfamiliar legal problems, junior lawyers tend to reason adaptively and hierarchically, drawing on contextual cues and engaging in iterative, recursive problem-solving.[46] By contrast, the AI's linear and exhaustive execution fails to reflect these adaptive reasoning processes. Instead, it presents a flat list of densely packed tasks that lack clear prioritization or contextual framing. As a result, junior lawyers using AI are implicitly directed to skip many of the key steps that they would otherwise undertake to fully understand the relevant legal and factual issues. Alternatively, while they may seek to partially replicate this process, doing so can plausibly be as demanding as performing the task from scratch, thereby diminishing the practical utility of AI assistance.

Additionally, at least some of the expertise and judgment of human lawyers resides in invisible moves or intuitive assessments, gut-feel judgments, and analogies from past experiences. These tacit insights are unrecorded and unacknowledged in AI-generated plans, resulting in incomplete or "shallow" outputs for novice lawyers. For instance, a lawyer may sense that a client's stated business goals are inconsistent with their risk tolerance, prompting them to reframe their advice accordingly. AI outputs, however, mechanically list steps without such contextual adaptation, forcing junior lawyers to mentally reinsert these unlogged insights to make sense of the production. This further exacerbates cognitive overload and reduces trust in AI systems.

III. Implications and Questions

This Part builds on the tentative conclusions advanced in Part II—namely, that junior lawyers may struggle to evaluate the quality and implications of work produced with the assistance of AI. It argues that this dynamic can help explain current patterns of AI use in legal practice. It also offers insights into how AI is likely to reshape the profession going forward if, as we expect, these trends apply more frequently to work completed by junior lawyers than senior lawyers. Subpart B then outlines directions for

---

[46] This intuition is captured well by Frazier and Rozenshtein's description of academic legal reasoning, which they suggest often required "engaging directly with primary sources, wrestling with contradictory authorities, and laboriously crafting arguments has long served not merely as a method of scholarship production but as a crucial mode of scholarly thinking itself. In other words, the value of scholarship is as much in the journey as it is in the destination. AI assistance may inadvertently short-circuit the intellectual development that occurs through sustained, unmediated engagement with legal materials." Alan Rozenshtein & Kevin Frazier, Large Language Scholarship: Generative AI in the Legal Academy (April 01, 2025). Minnesota Legal Studies Research Paper No. 25-26, Available at SSRN: https://ssrn.com/abstract=5200768 or http://dx.doi.org/10.2139/ssrn.5200768

future research, including several potential empirical strategies for more precisely investigating when, why, and the extent to which lawyers struggle to assess the quality and implications of AI-assisted legal work.

A. Potential Implications of Junor Lawyers' Difficulty Assessing AI-Generated Content

If, as we argue in Part II, lawyers can struggle to assess and deeply understand the work they produce with the assistance of AI, this helps explain the legal profession's reluctance to fully embrace AI despite its potential to enhance the speed and quality of legal work. Although AI significantly improves lawyers' productivity on average, it also introduces the risk of undermining the quality of legal work in ways that are difficult to detect and correct. For a risk-averse profession, this potential for undetected error itself a powerful deterrent to widespread AI use, particularly among junior attorneys.

Although our empirical results focus exclusively on upper-level law students, it remains unclear whether they generalize to more experienced lawyers. Still, even if the findings apply only to relatively inexperienced lawyers, they help illuminate the profession's limited uptake of AI. Junior lawyers—who are typically more comfortable with new technologies and less bound by established routines—would seem the most likely adopters of AI tools. Moreover, many of the legal tasks that have been most deeply integrated with AI, such as legal research and document review, have historically been disproportionately performed by junior attorneys.

Beyond explaining current patterns of adoption, this assessment also has important implications for the future of legal practice. Most practically, it suggests that lawyers should be encouraged to use AI in specific targeted ways that are most likely to preserve the cognitive benefits of grappling with complex and unfamiliar legal problems. For instance, junior lawyers might be instructed to use AI to generate overviews of broad legal doctrines, suggest additional sources to cite or review, assist with Bluebooking,[47] improve writing style, or summarize material that the junior lawyer has already independently analyzed and come to deeply understand.[48] By contrast, law firms and senior lawyers might reasonably discourage AI use by junior

[47] See Matthew Dahl, Bye-bye, Bluebook? Automating Legal Procedure with Large Language Models, https://arxiv.org/abs/2505.02763.

[48] Cf. Elliott Ash, Aniket Kesari, Suresh Naidu, Lena Song, and Dominik Stammbach. 2024. Translating Legalese: Enhancing Public Understanding of Court Opinions with Legal Summarizers. In Proceedings of the Symposium on Computer Science and Law, CSLAW'24, pages 136–157, New York, NY, USA. Association for Computing Machinery.

lawyers for more demanding tasks that require independent legal reasoning or complex factual synthesis, such as answering complex research questions, marking up nuanced deal documents, or formulating legal arguments.

Unfortunately, drawing these distinctions in practice is challenging. For example, the line between high-level and in-depth legal research is often blurry. Similarly, the process of refining legal writing can itself spark insights and deepen understanding—benefits that may be lost if AI is used too early or too extensively in the drafting process. These ambiguities suggest that while AI can indeed improve the efficiency of lawyers' work, mitigating the associated risks to quality will require ongoing attention and judgment. They also imply that AI is unlikely to fundamentally transform junior lawyers' ability to perform complex legal tasks efficiently—at least not without significant safeguards and oversight, or significant risks of decreased quality.

Second, these conclusions suggest that AI may have greater transformative potential when deployed by experienced lawyers, who are likely to be comparatively well equipped to rapidly and accurately evaluate the quality of legal output and to appreciate its implications and limitations. Of course, the extent to which this evaluative capacity exists for senior lawyers will naturally vary across tasks, individuals, and contexts. However, we suspect that in a substantial number of cases, experienced lawyers possess sufficiently well-developed mental models to bypass the more exploratory and unstructured reasoning processes that less experienced lawyers often engage in when addressing novel or complex issues. As a result, experienced lawyers may be able to leverage AI across a broader range of tasks with relatively low risk of missing errors or misapplying its output—particularly when working within their established domains of expertise.

This conclusion stands in partial tension with empirical findings that AI tools tend to yield disproportionate performance gains for less experienced users, including law students.[49] However, this apparent contradiction can be reconciled: while AI may disproportionately enhance the work of novices, it also introduces distinct risks of degrading the quality of their work in ways that may be difficult to detect. By contrast, while AI is less likely to significantly enhance the substantive work of expert lawyers, it also poses less risk of degrading quality when applied to tasks and scenarios within their established areas of expertise. In these cases, the efficiency gains experts can achieve by using AI in familiar domains carry little downside risk. This dynamic is illustrated by a prominent UK judge who has publicly described his use of AI to draft opinions: he uses AI to generate paragraph-length summaries of legal doctrines he already knows well, reviewing the output

---

[49] See Choi & Schwarcz, supra.

before incorporating it into his opinions. Because he is deeply familiar with the underlying law, this workflow yields time savings without compromising quality.[50]

To the extent that more senior attorneys disproportionately benefit from AI compared to their junior counterparts, these dynamics may significantly reshape traditional law firm business models. Historically, many firms have operated under a pyramid structure, relying on large cohorts of junior lawyers supporting a smaller number of senior associates and partners. But if AI primarily enhances the efficiency of expert lawyers, firms may increasingly shift toward leaner models that employ fewer junior attorneys while placing greater emphasis on recruiting top junior talent with the potential to grow into future experts.

Of course, these predictions are highly speculative and raise numerous complications that we cannot fully address here. For example, it is far from clear whether law firms will continue investing in the development of junior lawyers in a world where most entry-level tasks can be performed as well—or better—by AI. If, as we suspect, junior lawyers primarily learn by engaging in realistic legal tasks, then many firms may opt to forego hiring them altogether. Alternatively, firms might choose to subsidize junior associates while they build the necessary skills to work effectively with AI. But doing so could prove economically unsustainable, especially if competitors undercut them by charging lower fees or poaching partially trained associates. Another possibility is that law schools will step in to ensure that graduates enter the profession with the capacity to critically assess and build upon AI-generated output. Yet even this is uncertain, as AI has the potential to undermining practice-oriented legal training. That is because many of writing assignments that are core elements of such practice-oriented training can no longer be assigned outside of artificially constrained settings, given the ease with which AI can complete them.

However these issues are ultimately resolved, the broader point remains: understanding the future of the legal profession requires a clear grasp of how attorneys' use of AI across can enhance, disrupt, or fundamentally reshape human workflows and reasoning processes. The data and hypotheses presented here represent an exploratory effort to address this question, but their limitations are evident. As a result, more targeted empirical research is essential. It is to that topic that we now turn.

---

[50] Gary E. Marchant, *AI in Robes: Courts, Judges, and Artificial Intelligence,* 50 OHIO N U L REV. 473 (2024).

B. Towards a Better Understanding of How AI Use Impacts Legal Reasoning

Although existing empirical research indicates that junior attorneys may have difficulties evaluating and deeply understanding work generated with the assistance of AI, this conclusion remains tentative and underexplored. The primary limitation is that existing studies were not explicitly designed to investigate this issue; rather, they provide only indirect or incidental evidence of this possibility. Given the significant practical implications of this possibility, there is a clear need for targeted empirical research specifically designed to evaluate it.

One promising empirical approach involves a sequential task experiment. In this design, upper-level law students or junior attorneys would complete two distinct but related legal assignments involving similar doctrines or factual scenarios. Participants would be randomly assigned to use AI on the first task only. Researchers could then compare performance on the second, AI-free task between the two groups. If those who used AI on the initial task perform worse on the subsequent assignment, it would suggest that AI use in early stages may short-circuit the cognitive engagement necessary for deeper legal reasoning. This experiment might also incorporate peer and self-assessment components. After completing each of the two tasks, participants would evaluate both their own work and that of peers. This extension would help determine whether AI use affects one's ability to critically assess legal reasoning. Researchers could then compare the evaluative accuracy and depth of participants who previously used AI with those who did not.

In addition to these experimental approaches, qualitative methods such as structured interviews and surveys can offer a valuable complement. These might target junior attorneys who regularly incorporate AI into their practice, probing their perceptions of how AI usage impacts their own ability to fully understand legal tasks or doctrines. Alternatively, qualitative studies of more experienced attorneys who use AI to help perform familiar tasks could provide additional insights into which types of tasks or practice areas are likely to be most dramatically impacted by AI in the near term.