



American College of Coverage Counsel 2024 Insurance Law Symposium University of Minnesota Law School Minneapolis, MN November 15, 2024

Session Panelists

Michael S. Levine

American College of Coverage Counsel

Hunton Andrews Kurth LLP Washington, DC 20037 mlevine@hunton.com

Daniel Schwarcz

University of Minnesota Fredrickson & Byron Prof. of Law Minneapolis, MN 55455 schwarcz@umn.edu

Kate Middleton

ECOLAB St. Paul, MN 55102 Kate.middleton@ecolab.com

Barbara O'Donnell

McAngus Goudelock & Courie Boston, MA 02109 Barbara.ODonnell@mgclaw.com

Michael S Levine, Alex Pappas, Iris Devriese, and Shivakumar Balasubramaniyan

Jun 12, 2024 🕔 20 min read

Summary

- Smart contracts serve to keep the law grounded in more modern, equitable contract doctrines that serve as a counterweight to classic contract theory.
- The smart contract offers tort-based considerations that may remove it from the exclusionary aspects of CGL and other traditional coverage.
- It may also redefine what it is to provide coverage for "property" as it becomes an indistinguishable hybrid of hardware, software, and data.



Thomas Vogel via Getty Images

In little more than one year, since the emergence of ChatGPT, artificial intelligence (AI) has ushered in a new era, transforming industries and redefining the way we approach

problem-solving. While the term "artificial intelligence" was coined in 1956, AI technology continues to advance, and it is crucial to evaluate its real-world impact and consider the challenges and opportunities it presents. This is particularly the case with insurance, given that it is insurance that will be looked to in the wake of mishaps involving AI.

If the November 2023 controversy over Sam Altman's status as the chief executive officer of OpenAI is any indication, AI has captured the world's attention, and for good reason. AI is predicted to grow "exponentially" over the next decade and may contribute up to 14.5 percent of gross domestic product in North America by 2030. 1 Very few, if any, industries, businesses, or people will go unaffected. The insurance industry, which is itself having a "Generative AI Moment," is no exception. 2 Indeed, as the consultancy McKinsey & Co. wrote, AI "will have a seismic impact" on all aspects of the insurance industry. 3

The first part of this two-part article will unpack several critical facets of that seismic shift, which is already reshaping the insurance world for insurers and policyholders alike, by delving into the intricate landscape of AI, focusing on its growing influence in the insurance industry and the legal challenges and opportunities that arise. We begin by discussing how AI is reshaping the insurance industry, and we include a survey of how AI is being deployed across different insurance functions such as underwriting and claims processing. We next analyze the use of AI in the context of litigation and how AI will affect the collection and introduction of evidence, issues that will ultimately affect the scope of liability insurance and associated coverage for defense costs.

In the second part of this two-part article, we will consider how the marketplace for Alspecific insurance might develop, including a discussion of the pros and cons of Al-specific insurance products, which continue to debut and evolve. If deployed thoughtfully, insurance can "help avoid legal issues of liability" and even "enhance the integration of Al into daily commercial routines while mitigating" potential downsides.

Together, this two-part article will provide guidance to members of the insurance bar about this rapidly evolving landscape where the fusion of legal and technological acumen will sculpt the future of the insurance business and insurance law, while creating opportunities for insurance practitioners. Indeed, this rapidly evolving discipline provides great promise for lawyers and other insurance professionals, in part because the new, rapidly developing issues provide a platform for insurance practitioners to make their mark.

The Role of AI in Commercial Insurance

Al is revolutionizing the insurance sector, with rising interest in Al algorithms to streamline processes, enhance customer experiences, and develop innovative insurance products.

From underwriting and claims processing to risk assessment, AI is reshaping the insurance landscape by providing data-driven insights and automating traditionally labor-intensive tasks. At its core, however, insurance is about clearly delineating what is covered from what is not. To do that requires clear and unambiguous wording. Definitions often must be supplied, particularly where technology and other concepts beyond the main are involved. Al is no exception. In fact, as we discuss later below and in greater depth in the second part of this two-part article, the failure to clearly define AI may lead to abject failure of the insurance product.

Types of AI. Broadly speaking, there are at least seven types of AI. Understanding which AI systems your company is running or your insurance is covering (or excluding) is fundamental to managing AI risk. Confounding even the clearest definitions and explanations, however, is the reality that many companies are not using just one type of AI or multiple types of AI in the same combinations. Complexity and technical inside baseball aside, knowing which systems are being used or insured is critically necessary to managing the AI risk.

- **1** *Reactive machines AI:* These are the simplest forms of AI systems that are purely reactive and can neither form memories nor use past experiences to inform current decisions. They are meant to perform specific tasks, and their behavior is entirely deterministic.
- 2 *Limited memory AI:* These AI systems can learn from historical data to make decisions. They can store past experiences or data for a brief time. An example of this is self-driving cars that observe other cars' speed and direction.
- **3** *Theory of mind AI:* This is a more advanced type of AI that can understand thoughts and emotions that affect human behavior. This AI system can interact socially. But it currently exists only in theory.
- **4** *Self-aware AI:* This is the final stage of AI development and it is currently hypothetical. Self-aware AI, which currently exists only in theory and science fiction, would be systems that have their own consciousness and self-awareness.
- **5** *Artificial narrow intelligence (ANI):* Also known as "weak AI," this type of AI is meant to perform a narrow task, such as voice recognition. These systems can only learn or be taught how to do specific tasks.
- **6** Artificial general intelligence (AGI): Also known as "strong AI," this type of AI refers to a system that possesses the ability to perform any intellectual task that a human can do. Such systems can understand, learn, adapt, and implement knowledge in a broad range of tasks.

7 *Artificial superintelligence (ASI):* This refers to a time when the capability of computers will surpass that of humans. ASI is currently a hypothetical concept often depicted in science fiction. It is proposed to have extraordinary cognitive capabilities, including the ability to understand and master any intellectual task that a human can do.

As insurance stakeholders work to derive a functional scope of coverage, definitions of AI will have to consider all types of AI. Failing to do so could lead to uncertainty of scope and ambiguity. Two existing definitions illustrate the dilemma. The first definition comes from the European Union's recently enacted Artificial Intelligence Act (EU AI Act). That regulation provides the following definition of AI:

Al system means a machine-based system designed to operate with varying levels of autonomy, that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments[.]⁵

From a functional standpoint, a definition like that used in the European Union's AI Act offers potential promise for insurance stakeholders looking to ensure a stable and predictable scope of coverage.

In contrast, one domestic insurer's recent attempt to define AI for purposes of an optional policy endorsement that seeks to exclude "content created or posted for any third party . . . created using generative artificial intelligence in performance of your services" ⁶ epitomizes the definition of circularity. The endorsement defines "generative artificial intelligence" to mean "content created through the use of any artificial intelligence application, tool, engine, or platform" ⁷ and thereby offers little guidance to its users.

Regardless of the definition deployed in a particular instrument, the question for insurance industry participants going forward should remain constant: how to define scope in a manner that achieves consistency and reasonable contractual certainty. The answer to this question can have wide-ranging, multibillion-dollar implications.

How AI is used in commercial insurance. Most insurers are focused on searching, summarizing policies, and synthesizing information to provide content and answer questions based on what AI has learned. There is also increased interest in decision support (not decision-making) in the underwriting process to assist underwriters. By analyzing vast and abstract sources of data and information and having the ability to detect patterns that might escape human cognition, underwriters can focus on the most

valuable risks. Likewise, claim handlers can use vast amounts of data to expedite the review of claims. But the use of AI also brings challenges, including allegations of discriminatory conduct, bias, data privacy concerns, and concerns over systemic inaccuracies without sufficient human oversight. Query, however, whether socially unacceptable outputs result from bias or simply objective analytics. Recent legislation tries to grapple with this dilemma.

1. Key technologies driving AI in insurance

- Machine learning: This technology enables computers to learn and improve from experience without being explicitly programmed. In insurance, it is used for risk assessment, fraud detection, and personalized policy pricing.
- Natural language processing (NLP): NLP allows computers to understand, interpret, and generate human language. In insurance, it is used for chatbots, claims processing, and customer service.
- Computer vision: This technology enables computers to interpret and understand the visual world. In insurance, it is used for tasks such as damage assessment in claims processing and risk assessment.
- Predictive analytics: This technology uses data, statistical algorithms, and machinelearning techniques to identify the likelihood of future outcomes based on historical data. In insurance, it is used for risk assessment, pricing, and claims prediction.

2. Impact on underwriting

Automation of routine tasks expedites decision-making, reduces operational costs, and allows underwriters to focus on complex aspects. Al's continuous learning enables dynamic risk assessment, crucial in a rapidly changing landscape. Sample use cases on underwriting include:

Risk Assessment: AI can improve the risk-assessment process by being trained on demographic data to better predict risk and provide underwriters with recommendations.

Intelligent Underwriting: Al can be used to identify critical documents, extract critical data in the submission process, and then feed just that critical information to the underwriter to help make quicker decisions.

Eligibility & Product Match: Al could be used to determine eligibility based on classifications and eligibility guidelines then suggest the best product match for the customer.

Social Media Sourcing: Al can be used to source social media to gather data around and confirm customers' business operations, social interactions and customer reviews.

Rating Errors: Al can generate notifications for underwriters when rating errors have been made, the impact, and the correction needed.

Policy Manuscript Generation: Al can generate basic policy manuscripts based on class codes or operations descriptions, or even personalize a manuscript based on exposure information.

Broker Messaging: Al can generate routine human-like communications in real time from underwriters to brokers when additional information is needed in assessing a risk.

While we believe that these examples represent a wide range of generative AI use cases in insurance underwriting, it is still a non-exhaustive list given the speed at which AI is advancing.

3. Implications of AI-driven risk assessment

a) Improved risk assessment

Today machines can aggregate and interpret data and can prioritize vulnerabilities, contextualize risk scoring, and measure exposures and countermeasures independently, resulting in more precise risk evaluations.

b) Automation of underwriting processes

We are also seeing increased opportunities to leverage AI and automate and streamline the data collection and analysis process, reducing the time and effort required for risk assessment. Using AI algorithms to analyze large volumes of data and identify patterns and trends, insurers are exploring ways to assess risk, improve efficiency, and reduce operational costs.

c) Impact on premium pricing

Al transforms premium pricing in insurance by enabling precise underwriting through data-driven insights. It facilitates dynamic pricing models that adapt to real-time risk factors, incorporates usage-based metrics (e.g., telematics in auto insurance), and detects and mitigates fraud. Al-driven predictive modeling anticipates future risks, allowing insurers to proactively adjust premiums. Customer segmentation and behavioral analytics enable personalized premium pricing, enhancing competitiveness, and customer satisfaction. Overall, AI improves accuracy, responsiveness, and customization in setting premiums, optimizing the balance between risk and pricing in the insurance industry.

Insurance Claims and Insurance Litigation

It takes little imagination to recognize the potential for AI to affect insurance underwriting, claim processing, and even the litigation of disputed claims. The use of AI in claims processing is no longer hypothetical, with multiple insurers already falling under attack for how AI is aiding their claims handling. The online insurer Lemonade has deployed its AI technology—AI Jim—to purportedly streamline and add efficiency to its claims process. ⁹ Yet, despite the advent and use of technologies like AI Jim, the use of AI in claims processing remains new. And because AI's use in claims processing is an unfamiliar legal area, there are not currently many fixed legal rules governing insurers' conduct in this space. ¹⁰ But one thing is clear now: For every potential benefit AI offers insurers in the claims process, corresponding legal risks must be considered. Indeed, only by taking a proactive approach that considers all the pertinent angles can relevant stakeholders avoid unwitting AI-generated pitfalls.

Al and insurance claims. Al's impact on claims processing is a two-sided coin. That is, Al can revolutionize claims processing, but it may also come at a substantial cost for both policyholders and insurers. Starting with the potential benefits, Al-driven claims processing could increase efficiency by automating various routine tasks, ranging from data collection and documentation analysis to fraud detection. Such automation may reduce the time required to process claims, enabling insurers to provide quicker responses to policyholders. Faster claims resolution may contribute to increased customer satisfaction and loyalty.

Depending on how AI claims technologies are deployed, insurers could also minimize human errors that have given rise to liability under state bad-faith statutes for inadequate or faulty claims handling. ¹¹ One reason is that AI systems, equipped with machine-learning algorithms, could analyze vast datasets with precision and thus potentially improve the reliability of claims processing while removing the risk of human-centric animus.

Despite these possible benefits, the deployment of AI in claims processing is not without potential drawbacks. As noted throughout this article, the potential for bias in AI algorithms is substantial. That is, if the training data used to develop these algorithms reflect historical biases, the AI systems may exacerbate or perpetuate these inequalities. Data privacy is, as detailed below, another critical risk associated with AI-driven claims processing. Further, as AI systems take on more decision-making roles in insurance,

questions arise about the transparency of these decisions and the accountability of algorithms. For example, there remains a real possibility that AI algorithms could be programed to reflexively deny claims or limit payouts despite contrary policy language and applicable background legal principles.

These benefits and drawbacks have only recently started to influence state insurance regulation. For example, as of April 30, 2024, 11 jurisdictions have adopted a model NAIC bulletin aimed at regulating the use of AI in the insurance industry. ¹² Four jurisdictions (California, Colorado, New York, and Texas) have also adopted insurance-specific regulations or guidance relative to AI. ¹³ State regulations have focused on avoiding discriminatory outcomes, among other things. ¹⁴ However, because government regulation in this area is in its relative infancy, only time will tell how much state-specific regulation will affect AI-driven claims handling.

Not only have state governments been calling out the risks of AI-driven claims handling so too have class action plaintiffs, as shown by recent lawsuits against health insurers like UnitedHealth and Cigna. ¹⁵ Take for example the UnitedHealth lawsuit pending in the U.S. District Court for the District of Minnesota. There, the estate of a deceased plaintiff has sued United Healthcare on behalf of a putative class of plaintiffs alleging that United illegally deployed AI "in place of real medical professionals to wrongfully deny elderly patients care owed to them . . . by overriding their treating physicians' determinations as to medically necessary care based on an AI model" that United allegedly knew had a "90% error rate." ¹⁶ Based on this overarching allegation and other supporting factual allegations, the plaintiffs alleged a breach of contract claim, a breach of the implied covenant of good faith and fair dealing claim, an unjust enrichment claim. and state law bad-faith claims. The lawsuit against Cigna involves similar allegations. ¹⁷ While these cases are in their early stages, the allegations themselves show how litigation over AI in claims processing might develop.

A duo of 2022 decisions—one from Washington and one from Delaware—confirms that Algenerated claims processing may give rise to legal liability. ¹⁸ In the Washington case, the Washington Court of Appeals held that a health insurer's practice of using a computer database to determine the reasonableness of a medical charge amounted to an unfair trade practice because the insurer did not undertake an individualized review. ¹⁹ But across the country in Delaware, the Delaware Supreme Court instead emphasized the reasonableness of fees rather than the process used to determine whether fees are reasonable. ²⁰ One takeaway from these cases is that insurers may have to justify not only their ultimate decision on a claim but also the process used to reach that decision.

A 2016 District of Arizona decision even confirms that challenges to technology-driven claims processing could get past the pleadings stage—and even summary judgment. In that case, a plaintiff alleged that an insurer was negligent and breached its duties by "improperly using . . . inadequate software" to deprive the insured of coverage under a homeowner's policy. ²¹ As to this negligence claim and theory, the court denied the insurer's motion for summary judgment, reasoning that "it may have been negligent for defendant to rely solely on its computer system to determine policy limits," among other features. ²² This case signals one potential restraint on insurers' conduct: Insurers are likely to be required to retain individualized human-centric review as part of their processes, no matter how good Al becomes in the near term.

While policyholders can state viable claims relative to AI technology in the insurance industry, a 2018 Eastern District of Pennsylvania case reaffirms that courts generally require a plaintiff to prove up specific flaws with a given computer-assisted technology.²³ In that case, an insured pursued a bad-faith claim against an insurer that used a computer model called Xactimate to calculate depreciation without "investigating the 'assumption models' Xactimate relies on."²⁴ The court rejected the insured's argument, stating that it did "not persuade th[e] Court."²⁵ The court reasoned that the Xactimate program was already an "industry standard computer program" and stated that the insured's argument would have been "stronger" if it involved "specific evidence" of how the Xactimate model was flawed.²⁶ The court also emphasized that generic complaints about assumptions were insufficient; the insured had to present "evidence that those assumption[s] [were] unreasonable...."²⁷

Taken together, the use of AI in claims processing brings with it many potential advantages and obstacles. As these issues are increasingly litigated, courts, policyholders, and insurers alike will want to monitor the rules that develop. The developing case law and the increased state-driven regulatory interest show a high degree of uncertainty about legal liabilities created using AI in the claims process. This new field also raises litigation-specific uncertainties, including under the Federal Rules of Civil Procedure and the Federal Rules of Evidence relative to both the discoverability and admissibility of AI-generated evidence. That is, as cases like *United Healthcare* and *Cigna* get past the pleadings stage (if they do), it will become essential for lawyers to consider how best to learn about and litigate relative to AI-generated evidence.

Litigating claims involving Al-generated evidence under the Federal Rules of Civil Procedure and the Federal Rules of Evidence. Electronic evidence is and has been essential in twentyfirst century legal proceedings.²⁸ For some time now, courts have been grappling with the discoverability and admissibility of text message evidence, mobile communications,

and social media posts, among other types of electronic evidence. ²⁹ Since at least a 2012 ruling from the U.S. District Court for the Southern District of New York, courts have permitted the use of machine-learning tools to help with e-discovery. ³⁰ And in 2016, the Wisconsin Supreme Court held that courts may consider predictive modeling when imposing a sentence, even though courts may not rely solely on predictive modeling for the sentence imposed. ³¹ While certain technological advances like these have been accepted by the courts, AI still represents a new frontier that will transform litigation generally and insurance litigation specifically.

One important question is how Al-generated evidence will be treated under the Federal Rules of Civil Procedure. In proceedings governed by the Federal Rules, discoverability is determined by Federal Rule of Civil Procedure 26, which provides that "[p]arties may obtain discovery regarding any nonprivileged matter that is relevant to any party's claim or defense and proportional to the needs of the case. . . ."³² Information "need not be admissible in evidence to be discoverable."³³ The U.S. Supreme Court has further cautioned that these rules should be applied broadly ³⁴ because the "[m]utual knowledge of all the relevant facts gathered by both parties is essential to proper litigation." ³⁵ This broad discovery standard is the standard against which Al-generated evidence will be judged. And because the standard for discoverability is so broad, courts are likely to at least allow some discovery relative to Al-generated content.

But as discussed below, vexing questions relate to whether AI-generated output is like testimony—and, if so, should those against whom the testimony is offered have a right to examine that evidence, thereby subjecting the generative algorithm and data to discovery —and whether a software application or algorithm is even to be considered "AI."

Other questions also remain about the reliability and authenticity of AI-generated content when courts are evaluating whether that evidence is admissible under the Federal Rules of Evidence. ³⁶ And the only way to identify whether AI-generated content is reliable or authentic is to allow discovery about it. It follows that courts are likely to allow at least some amount of discovery relative to AI-generated content. The tougher questions are the parameters of such discovery.

1. *People v. Wakefield* (N.Y. 2022): Addressing the scope and practicalities of AI in the courtroom

People v. Wakefield discusses the use of AI in forensic analysis, specifically the use of the TrueAllele system to interpret DNA evidence.
The court's decision questions the reliability of AI in a legal context and the potential implications for defendants' rights. But

the case does not provide a definitive answer on whether defendants should be granted access to proprietary source code to challenge the reliability of AI systems.

The primary issue in *Wakefield* is the admissibility of the TrueAllele software's results under the *Frye* standard. The court found that the software was reliable and admissible, but the case raises other concerns about the use of AI in the criminal justice system. The defendant keenly argued that the AI-generated output was like an expert offering opinion testimony; and, thus, he was denied his right to confront witnesses because he was not given access to the software's source code. The court explained:

Defendant further argues that the trial court's denial of his request for the source code so that an expert could review it was a violation of his constitutional right to confrontation. The Sixth Amendment Confrontation Clause provides that, "[i]n all criminal prosecutions, the accused shall enjoy the right . . . to be confronted with the witnesses against [them]" (*Crawford v Washington*, 541 U.S. 36, 42 [2004]).

. . . .

Although a computer cannot be cross-examined, as Dr. Perlin explained, the computer does the work, not the humans, and TrueAllele's artificial intelligence provided "testimonial" statements against defendant as surely as any human on the stand. ³⁸

The court did not rule definitively on these issues, but it did acknowledge that the use of AI in the courtroom raises profound questions that will likely plague courts for years to come, even characterizing that breadth as potentially destabilizing:

The use of artificial intelligence within our system of justice presents challenging questions and may destabilize our established notions of the dividing line between opinion and uncontestable fact (*see e.g.*, Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L Rev 54, 62–82 [2019]; Andrea Roth, *Machine Testimony*, 126 Yale LJ 1972, 2021–2022 [2017]). Courts across the country will decide how our federal and state constitutions may be interpreted in light of continued technological advances and their application in the courtroom. ³⁹

2. *People v. Burrus* (N.Y. Sup. Ct. Sept. 8, 2023): Discussing whether a software application or algorithm is, itself, AI

*People v. Burrus*⁴⁰ is of interest because it discusses the definition and application of AI in the evidentiary context. Like *Wakefield*, *Burrus* also speaks from the perspective of forensic DNA analysis. The decision highlights the importance of clearly defining AI and how a failure to do so could lead to ambiguity. In *Burrus*, an expert in forensic biology testified that the FST (DNA analytics) software did not fit a particular definition of AI because that platform does not use machine learning, neuronets, or decision trees. The same expert later testified, however, that the FST platform *did* qualify as AI when defining AI more broadly to include automated decision-making systems.

What Does the Future Hold for AI in the Courtroom?

Because the use of Al-generated content in court proceedings is in its infancy, it is too early to tell how courts will evaluate the newest discovery challenges posed by Al. Early indications are still that Al will transform discovery rules, including under Rule 26, which generally dictates what is and is not discoverable. For instance, according to William Eskridge Jr., a professor of public law at Yale Law School, Rule 26(b)'s proportionality requirement may be challenged by Al. ⁴¹ One reason is that Al technologies may allow lawyers to review more documents at a lower cost, which may reshape current notions of proportionality. Other commentators have noted that Al technologies may also require greater up-front discussion to make sure that all parties and courts are on the same page as the case proceeds. ⁴²

Because the standards for discoverability are laxer than the standard for admissibility, more complicated questions relate to how insurance lawyers and litigators can approach evidentiary issues under the Federal Rules of Evidence. The largest AI-specific challenges are likely to relate to the authenticity and reliability of AI-generated content and testimony, rather than threshold showings of relevance. Even though the relevance standard is more stringent under the Rules of Evidence than the Rules of Civil Procedure, the required threshold showing is still not incredibly high. ⁴³

Although the relevance threshold is moderately low, Federal Rule of Evidence 403 still provides a colorable basis to exclude certain Al-generated evidence. Federal Rule 403 provides that the "court may exclude relevant evidence if its probative value is substantially outweighed by a danger of one or more of the following: unfair prejudice, confusing the issues, misleading the jury, undue delay, wasting time, or needlessly presenting cumulative evidence." ⁴⁴ While courts generally interpret Rule 403 in favor of admissibility, Rule 403 still provides potentially strong grounds for a court to deny the admission of Al-generated evidence. ⁴⁵ The reason is that AI technology may cause unfair prejudice, confuse the issues, or confuse a jury. And judges may not be ideally positioned to determine whether a

jury can be misled by AI evidence without first understanding how the technology works. Likewise, judges may be unable to assess the likelihood of jury confusion without understanding whether the AI being considered in a case is valid and reliable. ⁴⁶ In this way, the Rule 403 analysis at least is partially dependent on the two most vexing AI-related evidentiary questions: authenticity and reliability.

Proving the authenticity and reliability of an AI technology may require counsel to do more legwork than would otherwise be required for more generally accepted or well-known technologies. ⁴⁷ For example, without training the court about the development and use of the AI, it will be very difficult for a court to determine the reliability or relevance of that evidence. ⁴⁸ Anticipating the need for greater explanation, trial judges may ask that the parties apprise the court early-on about whether they intend to offer AI evidence, perhaps requesting briefing or limited discovery to inform the issues. ⁴⁹ The greater complexity of AI systems may also diminish the frequency of contemporaneous evidentiary rulings in favor of up-front and thorough judicial processes and procedures for determining the admissibility of AI-generated evidence.

Apart from Rule of Evidence 403, the authentication of Al-generated evidence raises questions under Rules 901(a) and 602. Rule 901(a) provides that "[t]o satisfy the requirement of authenticating . . . an item of evidence, the proponent must produce evidence sufficient to support a finding that the item is what the proponent claims it is." ⁵⁰ Rule 602 in turn establishes the need for an authenticating witness, which arguably means that such witness must know about how the AI technology functions to authenticate it. ⁵¹ Because of the complexity and novelty of certain AI technologies, multiple witnesses may be required. ⁵² One solution may be the use of an expert to authenticate the AI technology, which would allow the witness to testify based on inputs received from others.

But expert testimony will not be without challenges. Expert testimony, as always, is subject to additional scrutiny under Federal Rules of Evidence 702 and 703 and *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, ⁵³ and *Kumho Tire Co. v. Carmichael* ⁵⁴ and their progeny. These rules and cases require that an expert witness provide reliable testimony based on sufficient facts or data that results from reliable principles and methods that have been reliably applied to the facts of the case. One of the reasons is that "[u]nreliable evidence has no tendency to prove or disprove facts that are of consequence to resolving a case or issue." ⁵⁵

Heeding the above rules, insurance practitioners should brush up on the Rules of Evidence and Civil Procedure. And even if battles over discoverability and admissibility are lost, the weight afforded to any AI evidence still is subject to question. That is, even if AI technology bypasses the gatekeeper, deficiencies and biases still present obstacles before the trier of fact. ⁵⁶

Stay tuned for the second part of this two-part article, which will be published in the next issue of Insurance Coverage.

Endnotes

- 1. Anat Lior, "Insuring AI: The Role of Insurance in Artificial Intelligence Regulation," 35 *Harv. J.L. & Tech.* 467 (2022).
- 2. Christopher Freese, "Leading Insurers Are Having a Generative Al Moment," Bos. Consulting Grp., Aug. 17, 2023.
- 3. Ramnath Balasubramanian et al., *Insurance 2030–The Impact of AI on the Future of Insurance*, (McKinsey & Co. Mar. 12, 2021).
- 4. Lior, "Insuring AI: The Role of Insurance in Artificial Intelligence Regulation," *supra*.
- 5. Artificial Intelligence Act, art. 3, Eur. Parl. Doc. P9_TA(2024)0138 (Mar. 13, 2024).
- 6. Philadelphia Consolidated Holding Corp., Musical Composition and Generative Artificial Intelligence Exclusion, Form PI-IT-036 (09/23).
- 7. Philadelphia Consolidated Holding Corp., Musical Composition and Generative Artificial Intelligence Exclusion, Form PI-IT-036, *supra*.
- Abraham Gross, "<u>Colo. Al Bias Law Brings Little Certainty For Insurance Sector</u>," *Law360*, May 23, 2024.
- 9. Ilkhan Ozsevim, "<u>Lemonade Sets World Record with 2-Second AI Insurance Claim</u>," *AI Mag.*, June 14, 2023.
- 10. While there are not yet many formal rules, the National Association of Insurance Commissioners (NAIC) issued a model bulletin on the use of AI systems in insurance. The bulletin is one step among many to create a comprehensive set of regulatory standards to ensure the responsible deployment of AI in the insurance industry. *See* NAIC Model Bulletin, <u>The Use of Artificial Intelligence Systems in Insurance</u> (Dec. 4, 2023).
- 11. *Cf. Carrol v. Allstate Ins. Co.*, 815 A.2d 119, 130 (Conn. 2003) (holding that the evidence supported a faulty human-driven investigation, in part because the operative people conducted a hasty, incomplete, and ill-motivated investigation).
- 12. NAIC, <u>Implementation of NAIC Model Bulletin, Use of Artificial Intelligence Systems by Insurers</u> (map) (Apr. 30, 2024).

- 13. NAIC, <u>Implementation of NAIC Model Bulletin, Use of Artificial Intelligence Systems by Insurers</u> (map), *supra*.
- 14. Daphne Zhang, "<u>Insurers' Al Use for Coverage Decisions Targeted by Blue States</u>," *Bloomberg L.*, Nov. 30, 2023.
- 15. Ken Alltucker, "<u>Is Your Health Insurer Using AI to Deny You Services? Lawsuit Says Errors</u> <u>Harmed Elders</u>," *USA Today*, Nov. 19, 2023; Richard Nieva, "<u>Cigna Sued Over Algorithm Allegedly</u> <u>Used to Deny Coverage to Hundreds of Thousands of Patients</u>," *Forbes*, July 24, 2023.
- 16. <u>Complaint</u> at 1, *Lokken v. UnitedHealth Grp., Inc.*, No. 0:23-cv-03514-WMW-DTS (D. Minn. Nov. 14, 2023), ECF No. 1.
- 17. Complaint, Kisting-Leung. v. Cigna Corp., No. 2:23-at-00698 (E.D. Cal. July 13, 2023).
- 18. Compare Schiff v. Liberty Mut. Fire Ins. Co., 520 P.3d 1085 (Wash. Ct. App. 2022), review granted, 526 P.3d 844 (Wash. 2023), with GEICO Gen. Ins. Co. v. Green, 276 A.3d 462 (Del. 2022).
- 19. Schiff, 520 P.3d at 1094-95.
- 20. GEICO, 276 A.3d at 462.
- 21. *Lewis v. Allstate Ins. Co.*, No. 3:15-cv-8074-HRH, 2016 WL 5408332, at *6–7 (D. Ariz. Sept. 28, 2016).
- 22. Lewis, 2016 WL 5408332, at *7.
- 23. *Sands v. State Farm Fire & Cas. Co.*, No. 5:17-cv-4160, 2018 WL 1693387, at *5 (E.D. Pa. Apr. 6, 2018).
- 24. Sands, 2018 WL 1693387, at *5.
- 25. Sands, 2018 WL 1693387, at *5.
- 26. Sands, 2018 WL 1693387, at *5.
- 27. Sands, 2018 WL 1693387, at *5.
- UNESCO, <u>How to Determine the Admissibility of Al-Generated Evidence in Courts?</u> (July 21, 2023; last updated July 26, 2023).
- 29. See UNESCO, How to Determine the Admissibility of AI-Generated Evidence in Courts?, supra.
- 30. See Moore v. Publicis Groupe, 287 F.R.D. 182 (S.D.N.Y. 2012).
- 31. See State v. Loomis, 881 N.W.2d 749 (Wis. 2016).

- 32. Fed. R. Civ. P. 26(b)(1).
- 33. Fed. R. Civ. P. 26(b)(1).
- 34. *Hickman v. Taylor*, 329 U.S. 495, 506 (1947) ("[D]iscovery provisions are to be applied as broadly and liberally as possible. . . .").
- 35. *Hickman*, 329 U.S. at 507.
- 36. Paul W. Grimm et al., "Artificial Intelligence as Evidence," 19 *Nw. J. Tech. & Intellectual Prop.* 9 (2021).
- 37. People v. Wakefield, 195 N.E.3d 19 (N.Y. 2022).
- 38. Wakefield, 195 N.E.3d at 23-24.
- 39. Wakefield, 195 N.E.3d at 24.
- 40. People v. Burrus, No. 817/2020, 2023 N.Y. Misc. LEXIS 5805 (N.Y. Sup. Ct. Sept. 8, 2023).
- 41. <u>Cassandre Coyer, "Generative AI and Federal Rules of Civil Procedure: Is It Meant To Be?</u>," *ALM*, Oct. 13, 2023.
- 42. See Coyer, "Generative AI and Federal Rules of Civil Procedure: Is It Meant To Be?," supra.
- 43. Federal Rule of Evidence 401 provides that evidence is relevant if "(a) it has any tendency to make a fact more or less probable than it would be without the evidence; and (b) the fact is of consequence in determining the action." This is generally a low bar. That is, under this standard, Al-generated content has at least some tendency to make a fact more or less probable.
- 44. Fed. R. Evid. 403.
- 45. Am. Ass'n for the Advancement of Sci. (AAAS), <u>Artificial Intelligence, Trustworthiness, and</u> <u>Litigation</u> (Sept. 2022).
- 46. AAAS, <u>Artificial Intelligence, Trustworthiness, and Litigation</u>, supra, at 12.
- 47. AAAS, Artificial Intelligence, Trustworthiness, and Litigation, supra, at 14–15.
- 48. AAAS, Artificial Intelligence, Trustworthiness, and Litigation, supra, at 13-14.
 - [48] AAAS, <u>Artificial Intelligence, Trustworthiness, and Litigation</u>, supra, at 13–14.
- 49. AAAS, <u>Artificial Intelligence, Trustworthiness, and Litigation</u>, supra, at 12.
- 50. Fed. R. Evid. 901.

51. Fed. R. Evid. 602.

- 52. AAAS, Artificial Intelligence, Trustworthiness, and Litigation, supra, at 13–14.
- 53. Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579 (1993).
- 54. Kumho Tire Co. v. Carmichael, 526 U.S. 137 (1999).
- 55. Sedona Conf., Commentary on ESI Evidence & Admissibility, Second Edition, 22 *Sedona Conf. J.* 83 (2021).
- 56. *See* Patrick W. Nutter, "Machine Learning Evidence: Admissibility and Weight," 21 *Univ. Pa. J. Const. L.* 919 (2019).

Authors

•••





•••

•••

Shivakumar Balasubramaniyan

Published by the American Bar Association ©2024. Reproduced with permission. All rights reserved. This information or any portion thereof may not be copied or disseminated in any form or by any means or stored in an electronic database or retrieval system without the express written consent of the American Bar Association.

ABA American Bar Association

https://www.americanbar.org/groups/litigation/resources/newsletters/insurance-coverage/artificial-intelligence-and-insurance-part-one/?login

Artificial Intelligence and Insurance—Part II, The Story Unfolds

Michael S Levine, Alex Pappas, Iris Devriese, and Shivakumar Balasubramaniyan

Aug 14, 2024 () 18 min read

Summary

- Al is poised to have a significant and lasting effect on the insurance industry.
- The use of AI algorithms can streamline processes, improve customer experiences, and facilitate the development of innovative insurance products.
- It also raises legal challenges, such as the potential for biased algorithms, data privacy concerns, and questions around the accountability and transparency of AI decisions.
- The evolving regulatory landscape and ongoing court cases will shape the future development of AI insurance.



breakermaximus via Getty Images

In Part I of this two-part article, we explored the transformative impact of artificial intelligence (AI) on the insurance industry. We delved into how AI is revolutionizing various insurance functions, including underwriting and claims handling. We also examined AI's influence on litigation, particularly how it affects the collection and introduction of evidence.

In Part II, we continue to unpack the key changes AI brings to the insurance landscape. We begin by exploring how AI-related risks can be transferred to insurers with a look at the insurability of AI-related risks, as well as an analysis of emerging insurance products that hold promise in this evolving field. We also take a closer look at AI and insurance from a risk management perspective through case studies on loss prevention, fraud detection, and predictive analytics. This examination highlights the potential for AI to reshape business. It also highlights how insurance products can help businesses mitigate risk as they continue to deploy AI across their various functions.

AI Risks Transferred to Insurers

Al brings not only benefits and improvements to the insurance industry and its processes. As more businesses across all industrial verticals are incorporating Al tools into their dayto-day operations and automating processes and decision-making, companies are increasingly worried about Al risk.

From privacy and data protection concerns to intellectual property infringement, the challenges are multifaceted. Among these, two critical risks—model bias and model underperformance because of data drift—loom large, casting shadows over the successful deployment of AI systems. Model bias refers to the inherent prejudices embedded in AI algorithms, leading to discriminatory outcomes. Data drift, the gradual evolution of input data over time, can cause AI models to underperform, affecting their accuracy and reliability.

These risks carry potentially significant financial implications for organizations. In fact, managing AI risks has been listed as the main barrier leaders face in scaling existing AI initiatives. 1 Insurance has enabled business ventures in countless previous instances—as Henry Ford famously said, "[w]ithout insurance we would have no skyscrapers. . . . " 2 Insurance could also prove to be the right vehicle to manage and transfer AI-related risks to support the safe adoption of AI by companies and society.

As we discuss below, traditional coverages do not fully protect against AI risks; they leave significant coverage gaps. We will touch on some of the available insurance coverages specifically for AI today.

Al as an insurable risk. For a risk to be insurable, it needs to be pure (resulting in a loss or no loss with no possibility of financial gain—to be contrasted with speculative risks like gambling), quantifiable (measurable in financial terms), and fortuitous (the insured event needs to occur by chance), and the corresponding losses need to be measurable. When analyzing the risk of model underperformance, it becomes clear that Al risks exhibit these elements of insurability.

What is model underperformance? Suppose an AI model classifies credit card transactions as fraudulent or not fraudulent. Further, suppose that the model correctly classifies these transactions 90 percent of the time on previously unseen test data—i.e., the AI model has an error rate of 10 percent as determined on test data. The performance of AI models (meaning the error rate) can fluctuate for various reasons; for example, data drift, which occurs at random and causes a spike in the error rate. As noted above, "data drift" refers to the unanticipated and gradual changes in the characteristics and distribution of the incoming data, introducing unexpected variations that can affect the performance and reliability of machine-learning models.

In our example, suppose the model correctly identifies only 80 percent of fraudulent transactions when actively used in the real world in a given month as the associations in the data change compared with what the model was exposed to in the test data (i.e., there is a data drift between test data and actual use case data in the given month). This data drift exposes users to double the amount of fraud claims than anticipated. More generally, in all scenarios where AI systems are crucial for operations, underperformance can result in losses, business interruptions, and decreased productivity. Transferring this statistical fluctuation risk on the error rate could be beneficial to many AI users as it creates financial certainty.

Insuring against AI underperformance falls within the domain of pure risk, as businesses seek coverage against the negative outcome of AI systems failing to meet predefined performance thresholds for reasons that cannot be fully mitigated technically. Establishing performance benchmarks and checking the AI's historical performance (or representative test performance) against those thresholds allows an estimation of the probability of underperformance. Insurers can then determine the financial impact, should these benchmarks not be met, which provides a basis for calculating premiums and payouts. As underperformance is often caused by data drift—by definition unanticipated—it aligns with the need for fortuitousness. As a result, all elements of insurable risks are met. ³

The insurability of other AI risks (e.g., intellectual property infringement, discrimination, liabilities) is less straightforward. For one, the legal environment around AI liability and intellectual property infringement is nascent, and the treatment of AI in courts is still very opaque, making it difficult for insurers to estimate potential losses and calculate corresponding premiums. Pending court cases and regulations will increase the

transparency of this risk. Furthermore, the quantifiability of other AI risks is more complex. This complexity can be simplified by tying those risks to a performance threshold (e.g., when measuring bias in the form of a fairness metric), framing the risks as performance risks, and then quantifying the risks as described above.

Uncertainty of traditional insurance coverage. Knowing that AI risks are insurable, our next focus is to determine whether and how existing policies protect against AI risks.

Given the widespread integration of AI, damages arising from AI-related incidents can manifest in various forms, including—as in the example above—financial losses and operational disruptions, but can also lead to data and privacy breaches, as well as legal liabilities including intellectual property infringement. The damages incurred may implicate a range of insurance policies, such as cyber insurance for data breaches, general liability insurance for physical harm caused by AI-based machinery, technology liability for negligence claims, and media liability for intellectual property infringement during AI model training.

Traditional insurance policies can offer coverage for certain AI-related losses, bridging some gaps but leaving significant areas unprotected. ⁴ General liability policies most likely cover AI-caused physical harm, but most liability policies exclude discrimination. Discrimination is covered in employment practices liability insurance but only for employment-related discrimination. As AI models are being used increasingly in various areas where laws against discrimination apply (e.g., healthcare, real estate, credit approvals), this could leave users uninsured against potential lawsuits (class actions).

Cyber insurance policies are effective against data privacy issues, but they may fall short in cases like the Samsung data leak, where employees inadvertently leaked source code while using ChatGPT to help with work. As the unauthorized disclosure of the code involved the insured's proprietary data, its cyber policy could refuse to pay its losses. Dependence on Al vendors' coverage poses challenges, especially considering the potential size of financial effects on businesses.

Technology liability policies are meant to cover third-party claims for negligence, yet the complexity of AI risks challenges their effectiveness. The "black box" nature of certain AI models complicates determining negligence, and uncertainties around applicable standards make these policies a primitive tool for AI risk protection.

Because AI risks are novel, existing policies will change over time to also encompass certain AI risks. For now, companies and AI providers face significant coverage gaps. The uncertainty around coverage makes it hard for companies and risk managers to fully assess their exposure to AI risks. This uncertainty is burdensome for the insured but could also expose insurers to unexpected risks that are not priced into the insurance policy. This

Artificial Intelligence and Insurance-Part II, The Story Unfolds

"silent AI" exposure will need to be thoroughly explored so as not to expose insurance portfolios to unexpected, significant, and potentially systematic risks and losses.

Available coverages today. To date, very few insurance policies consider AI risks. Few insurers are openly addressing or making public statements about the risks associated with AI, bringing up concerns about the industry's apparent reticence in acknowledging and mitigating potential challenges in this rapidly evolving technological landscape. A notable exception is Munich Re, a leading global provider of reinsurance, insurance, and insurance-related risk solutions. Munich Re has been insuring AI risks since 2018 and has emerged as a pioneer in providing insurance solutions to mitigate the financial ramifications associated with AI underperformance and other AI risks.

Underperformance: Munich Re's third-party flagship product protecting AI providers against model underperformance is called aiSure for performance guarantees. This insurance product allows AI providers to support the quality of their models, assuring customers of their AI tools' reliability by providing them with performance guarantees. Suppose an AI vendor would like to promise its customers a specific accuracy level, such as 90 percent, in fraudulent transaction detection. When the AI falls short of this commitment, Munich Re provides financial restitution aligned with the losses suffered. This insurance-backed performance guarantee aims to instill confidence in AI with Munich Re's financial stability, ensuring effective mitigation of risks associated with AI model underperformance.

In addition, Munich Re provides aiSure for its own AI models, addressing the needs of businesses implementing self-built ("home-grown") AI solutions. Suppose a car manufacturer is relying on AI to identify the need for paint jobs via cameras as part of the car production process. Munich Re's aiSure ensures protection against an AI error rate drifting beyond a predetermined threshold, which could otherwise leave the business unprotected against potential recalls and losses arising out of business interruption. This insurance solution enables enterprises to integrate AI models into critical operational tasks without undue concerns about potential underperformance.

Legal liabilities: While insurance solutions for AI underperformance are beginning to emerge, legal liabilities stemming from AI-related risks present an even more complex landscape. The intricate legal situation surrounding AI models and the evolving nature of court outcomes contribute to a scarcity of insurance products covering legal liabilities that are thus far not covered by traditional insurance solutions. As the responsibilities and legal implications of AI model failures remain largely uncharted territory, the lack of established risk-transfer solutions adds a layer of uncertainty.

Recognizing the need for comprehensive risk mitigation, Munich Re is now developing insurance products tailored to address legal liabilities associated with AI. Among these

offerings is aiSure for discrimination risks, a first insurance solution designed to safeguard businesses against damages and financial losses arising from lawsuits that allege AI-made decisions result in discrimination against protected groups.

The AI landscape is constantly evolving, and a budding of awareness of AI risks is growing into implementation concerns. With a jump in the number of businesses incurring financial losses due to AI failure, existing insurance policies are expected to change over time to affirmatively incorporate or specifically exclude many of the existing AI risks. Tailored insurance policies and endorsements that deal with specific AI risks are expected to multiply in the meantime, as users and providers become more aware of their exposure and as the legal landscape clarifies.

Outlook on the Developing AI Insurance Market

We offer an outlook on potential future market developments by sharing ideas about insuring generative AI and drawing parallels between the young cyber insurance market and the rising AI insurance market.

Challenges of insuring GenAl. Generative AI (GenAI) represents a significant evolution from traditional AI. While conventional AI models are designed for specific tasks, GenAI, exemplified by models like GPT-4 and Bard, can generate novel content—text, images, and more. This generative capability introduces new and unique risks, such as the potential for hallucinations, intellectual property infringement, the spread of false information, and the generation of harmful content. Unlike conventional AI, GenAI operates in an unsupervised or semi-supervised manner, responding with a degree of "creativity."

This creativity brings subjectivity and complexity to evaluating GenAl's outputs, making the risks associated with GenAl distinct and challenging. The difficulty arises in defining concrete thresholds for underperformance, as GenAl's outputs, such as hallucinations or false information, may not have a clear, objective benchmark. Testing regimes must be tailored to specific tasks, and the evaluation process involves considerations like the model's input space, clear definitions of undesired outputs, and the continuous monitoring required to capture performance changes over time. In addition, the updating of foundation models further complicates the underwriting process, requiring higher standards of monitoring and adaptation.

Munich Re outlines a framework ⁵ for insuring different risks associated with GenAI. Risks like hallucinations, false information, and harmful content could be insured by developing a model evaluation pipeline in collaboration with GenAI providers. The insurance would be based on defined performance metrics and thresholds, with a focus on specific tasks and a comprehensive testing regime. For addressing model bias and fairness, Munich Re proposes determining and agreeing on fairness metrics aligned with the application's goals. The evaluation involves defining thresholds and assessing the trade-off between fairness and accuracy. Munich Re also delves into the challenges of insuring against intellectual property and privacy violations, proposing methods like using narrow definitions agreed on by both insurer and insured and leveraging training techniques for quantifiable risks. But many risks, including environmental impacts, are still under exploration, and Munich Re plans to adapt its risk transfer solutions as the risk landscape and demand for protection become clearer.

Al insurance market and cyber insurance market. Much like the internet, Al has received wide adoption across nearly all corporate functions and all industries. For an outlook on Al insurance, analyzing the treatment of cyber risks and the corresponding insurance provides valuable insights into the potential future development of the Al risk market.

Much as the rise of cyber risks in the late 1990s prompted insurers to explore new territories, the surge in AI usage will soon become a focal point for emerging risk teams. AI insurance is in its early stages, akin to the initial forays into cyber insurance. The first cyber policies written focused on specific loss scenarios and were tailor-made with a strong technology focus. This seems to be the stage of AI insurance to date: To navigate the complexities of AI adoption, specific risks are addressed through tailor-made policies. Munich Re's underperformance insurance validates this theory with risk assessment focusing on the robustness of each individual model, premiums dependent on specific performance data, and payout structures developed case by case.

As businesses grapple with the transformative potential of AI, insurers will start developing coverage to manage AI-related liabilities. When losses from cyber incidents started spiking, risk managers, brokers, and insurers started thinking about cyber risks in a more systematic and strategic way. An increase in AI-related losses seems to be on the horizon, considering the recent uptick in intellectual property lawsuits, lawsuits against healthcare companies using AI, and an increased interest shown by regulatory agencies in not tolerating discrimination from AI models.

Regulatory landscapes, exemplified by the European Union's General Data Protection Regulation for cyber risks, play a pivotal role. Similarly, AI regulation will likely spur businesses to follow evolving guidelines and adopt responsible AI initiatives, creating a parallel with the regulatory journey in cyber insurance. Once these regulatory cyber landscapes were more clearly defined, markets started navigating compliance phases, developing standardized processes aligning with regulatory norms. This shift simplified underwriting and marked a transition toward an informed, standardized market practice, echoing the journey of other established insurance sectors. The ultimate vision is a mature AI insurance market, marked by standardized practices and structured pricing—akin to the evolution witnessed in cyber insurance.

Risk Management: AI-Generated Legal Risks

Al is changing the way businesses operate, and with those changes come many known and unknown—legal risks. ⁶ Among other things, Al might

- expose companies to additional cybersecurity and data privacy risks;
- give rise to product liability claims if AI-enabled products generate faulty (or even dangerous) outputs;
- create fiduciary liabilities for directors, officers, and managers who greenlight or fail to oversee AI deployment;
- result in intellectual property infringement;
- facilitate unwitting discrimination through algorithmic bias; or
- compel newly displaced employees to sabotage their former employers.

Because AI is largely novel, complex, and unregulated, AI may very well also generate unforeseen—and unintended—consequences. With this uncertainty, businesses face a panoply of risks that they may not fully understand or appreciate. Businesses should work to get ahead of these risks now before they face exposure later. In doing so, they should think about insurance and risk management issues early and often.

Just as no two businesses are the same, no two businesses have the same legal risk profile when it comes to AI. Potential legal liabilities turn on many factors, including the industry in which a business operates, the products or services it sells, and the corporate form it adopts, among other factors. ⁸ Together, these differences highlight an essential bottom line: Risk managers and insurance professionals must analyze the business fully to determine its risk profiles and exposures, which will differ even from others in the same field.

The three case studies below—which focus on the use of AI to address retail shrinkage, prevent fraud, or improve operational efficiencies—highlight the importance of businessand industry-specific AI audits.

Case study 1: Al-driven loss prevention. Retailers are increasingly deploying Al solutions to avoid losses in the form of theft or shrinkage, which has been growing rapidly in recent https://www.americanbar.org/groups/litigation/resources/newsletters/insurance-coverage/artificial-intelligence-and-insurance-part-two/?login

years. ⁹ The uptick in retail theft has cost retailers billions and even threatened shoppers and employees. ¹⁰ While AI offers a promising solution to retail shrinkage, it is not without unique legal risks.

Tangibly, retailers are using AI to complement existing anti-theft technologies. One of the goals of this AI-assisted technology is to catch thieves before they act. AI-assisted cameras that analyze images and detect suspicious activity are an example. AI cameras can not only monitor people in stores but can also monitor shelves, display cases, checkout lanes, and other areas in the store to detect theft before it occurs. This AI technology, together with a related suite of AI-enhanced technologies like acousto-magnetic electronic article surveillance systems and radio frequency identification systems, could be transformative for retail businesses seeking to minimize the rates of retail theft. ¹¹ As one commentator noted, intelligence-led loss prevention may not only thwart theft but also increase brand loyalty by using data garnered from surveillance activities to better understand specific customers' shopping habits. ¹²

Despite its promise, this technology brings with it many potential—and unique—legal risks. One prominent example is the potential for lawsuits against retailers by customers alleging civil rights violations through false accusations of shoplifting occasioned by AI technologies. That is, AI-driven loss-prevention technology may cause certain individuals to be singled out based on a protected characteristic like their race, sex, or age. Allegations of this sort even if untrue—might be very damaging for businesses, and not only because such allegations are prone to high levels of publicity that can cause large financial losses. Exposure is also heightened by the specter of class action litigation.

Preventive steps include, for example, the adoption of stringent loss-prevention policies and employee training programs. But while an ounce of prevention can often be worth a pound of cure, some risk of this type is likely to materialize in any event. Businesses should thus consider how their risk management and insurance programs can prevent (or minimize) any attendant financial exposure.

Another potentially unique risk exposure associated with AI-driven loss-prevention technology involves privacy-related concerns. Businesses may face lawsuits alleging that they violated customers' privacy-related rights. One example of a potential exposure relates to the Illinois Biometric Information Privacy Act (often called BIPA), which regulates the collection, use, and handling of biometric information and identifiers by private entities. BIPA—and other state-specific statutes, including Texas and Washington statutes —confirms that AI risk profiles are likely to vary, based on not only the specific use of AI but where that AI is used. In sum, AI offers considerable promise for retailers seeking to minimize retail shrinkage. But AI's promise is not without risk. The risks unique to AI-related shrinkage technologies underscore why risk managers and insurance professionals must analyze the business expansively to determine unique risk profiles and exposures. Indeed, only by thoughtfully considering all the various benefits and drawbacks can the full array of legal risks be addressed.

Case study 2: Al-driven fraud detection. Enhanced fraud detection is another way Al can benefit businesses' bottom lines. Like retail shrinkage, fraud is incredibly costly to individual firms and the broader economy. The recent "tidal wave of pandemic fraud" is just one example.
¹³ There, financial institutions were, on average, fleeced out of more than \$500,000 each within a year. By 2027, estimates are that fraud losses are likely to surpass \$40 billion.
¹⁴ Here too, Al-assisted technologies offer promise for businesses, including financial institutions, healthcare organizations, and even governments.

The promise of AI in fraud detection involves marked improvements to legacy technologies. Financial institutions, for example, have been using some form of technology-assisted fraud detection for decades. ¹⁵ But the success rate for these legacy technologies is low. Today's machine-learning systems and AI enhancements offer considerable promise for organizations of all types looking to improve upon legacy technologies. ¹⁶ They not only can better identify fraud before it happens; they can also reduce the number of false alerts associated with prior fraud-prevention systems. ¹⁷ It is therefore no surprise that banks and other institutions are increasingly looking to AI-driven tools as a potential solution. ¹⁸

Just as using AI to address retail shrinkage brings with it new risks, so too does using AI to address fraud. Indeed, as in other areas, the legal risks occasioned by AI in fraud detection will depend on exactly how AI is deployed and in what industry. ¹⁹ For example, healthcare organizations may use AI to detect fraud in medical billing, which could give rise to unique potential liabilities not faced in other arenas. ²⁰ AI also offers promise in reducing government-facing fraud by identifying gaps in gargantuan federal and state budgets. ²¹ But in the healthcare and government contexts, for example, there are unique risks, including under the federal False Claims Act. ²² Also unique to the fraud-prevention context, AI technologies may even unwittingly cause more fraud than at baseline levels. ²³

As these risks show, the use of AI in fraud detection presents a nuanced legal landscape that requires careful consideration as technology continues to evolve. While AI systems bring substantial advantages to the identification and prevent of fraud, they also raise

Artificial Intelligence and Insurance-Part II, The Story Unfolds

concerns that risk management and insurance professionals should consider. In the end, only by navigating this domain thoughtfully can organizations harness the benefit of Aldriven fraud detection while mitigating associated legal risks.

Case study 3: Predictive analytics. Al also offers promise for businesses seeking to improve their operations through predictive analytics. The term "predictive analytics" generally refers to the use of statistical algorithms to forecast future trends, enabling businesses to optimize inventory, improve delivery times, and potentially reduce costs. ²⁴ When using predictive analytics along with AI, companies may be able to identify even more insights that benefit their businesses. ²⁵ Companies that deploy AI in this way also face unique legal risks—just like retailers using AI for loss prevention and organizations using AI to decrease fraud. One key category of unique risks involves corporate litigation risks.

Suppose that predictive analytics harm a corporation's bottom line. That corporation—and its directors and officers—may face lawsuits alleging that they breached their fiduciary duties. These suits might take many forms, whether that be direct lawsuits, derivative lawsuits, or class action lawsuits. Before such a lawsuit is filed, corporations may face demands that they produce books and records under the Delaware books and records statute (i.e., Delaware General Corporation Law section 220) or other states' analogues. Businesses may also choose to disclose how they have used AI-driven predictive analytics to improve their business. In doing so, they face potential exposure under federal and state securities laws for the quality, content, and scope of those disclosures. None of these potential risks are static. They are all unique to exactly how a business is using predictive analytics to improve its operations.

Two specific and often discussed flaws highlight how predictive analytics can result in these types of corporate lawsuits. Predictive analytics may cause errors attributable to historical bias or otherwise faulty data inputs. ²⁶ In other words, because predictive analytics rely on historical data, they may produce faulty forward-looking outputs because of the inherent reliance on backward-looking data. Similarly, predictive analytics might otherwise include faulty data inputs that can harm a business's bottom line.

Another wrinkle is that corporate law could also develop such that corporations can be sued for *not* using AI. While this novel argument has not yet been tested in court, the argument would be that corporate law requires the use of AI because of its superior information-processing capabilities and the legal requirement that directors act on an informed basis.²⁷ As this example shows, the legal effect of AI is still being tested, which is yet another feature that businesses may want to consider as they contemplate their own unique AI risk profile.

All told, businesses should not reflexively assume that Al-driven business improvements are risk-free. Risks of all types abound, including corporate-law-specific risks that risk managers and insurance professionals would be wise to consider.

As the preceding case studies highlight, no two businesses are likely to face the same set of AI-generated legal risks. These differences highlight why businesses must consider AI risk holistically and conduct AI-specific audits of their particular business practices. Indeed, because insurance products and other risk management tools are often developed relative to specific risks, only by first understanding risks can those risks be adequately mitigated.

Conclusion

In conclusion, AI is poised to have a significant and lasting effect on the insurance industry. The use of AI algorithms can streamline processes, improve customer experiences, and facilitate the development of innovative insurance products. But it also raises legal challenges, such as the potential for biased algorithms, data privacy concerns, and questions around the accountability and transparency of AI decisions. Despite these challenges, with thoughtful risk management and the development of tailored insurance products, AI can offer substantial benefits to the insurance industry while mitigating potential risks. The evolving regulatory landscape and ongoing court cases will shape the future development of AI insurance and the legal frameworks surrounding AI in the insurance sector. ²⁸

Endnotes

- Stanford Univ. Inst. for Human-Centered Artificial Intelligence, <u>Artificial Intelligence Index 2023 Report</u> 214 (Apr. 2023).
- 2. Joaquin Hernandez Jimenez, "Insurance, the Necessary Driver for the Development of Society," *MAPFRE*, May 12, 2023.
- 3. It should be noted that insurers need to make sure that underperformance resulting from a data drift is not covered when it comes from business decisions, e.g., entering a new market without updating the AI model used.
- 4. Joe Quinn, "Emergent AI and Potential Implications for Cyber, Tech & Media Coverages," WTW, Nov. 9, 2023.
- 5. Munich Re, *Insure AI* (last visited Jan. 8, 2024).
- 6. Michael S. Levine et al., "Introducing The Hunton Policyholder's Guide to Artificial Intelligence," Hunton Ins. Recovery Blog, Sept.18, 2023.
- 7. Levine et al., "Introducing The Hunton Policyholder's Guide to Artificial Intelligence," supra.

- 8. Michael S. Levine et al., "<u>The Hunton Policyholder's Guide to Artificial Intelligence: The Importance of</u> <u>Auditing AI Risk</u>," *Hunton Ins. Recovery Blog*, Nov. 20, 2023.
- 9. Bob Woods, "<u>How America's Biggest Retailers Plan to Use Technology to Catch Organized Retail Theft</u>," *CNBC*, July 29, 2023.
- 10. See Woods, "How America's Biggest Retailers Plan to Use Technology to Catch Organized Retail Theft," supra.
- James Stark, "<u>Addressing Retail Theft with an Intelligent Loss Prevention Approach</u>," Sec. Mag., Aug. 11, 2023; Craig Szklany, "<u>AI Is Transforming Loss Prevention—But It's Not the Only Tool You Need</u>," Retail Info Sys., Oct. 31, 2023.
- 12. Bjoern Petersen, "Logs, Not Locks: How Intelligence-Led Loss Prevention Stops Crime While Building Loyalty," Forbes, May 17, 2022.
- Benjamin Chou, "<u>The Digital Sentry: How AI Will Revolutionize Financial Fraud Investigation</u>," Forbes, July 3, 2023.
- 14. Tatiana Walk-Morris, "Fraud Losses to Surpass \$40B by 2027," Payments Dive, June 16, 2023.
- 15. Lisa Quest et al., "The Risks and Benefits of Using AI to Detect Crime," Harv. Bus. Rev., Aug. 9, 2018.
- 16. See Quest et al., "The Risks and Benefits of Using AI to Detect Crime," supra.
- 17. See Quest et al., "The Risks and Benefits of Using AI to Detect Crime," supra.
- 18. Evan Weinberger, "Bank Regulators Look to Existing Tools to Police AI Advances," Bloomberg, July 10, 2023.
- 19. Yang Bao et al., "<u>Artificial Intelligence and Fraud Detection</u>," *in 1 Innovative Technology at the Interface of Finance and Operations* 223 (Volodymyr Babich et al. eds., 2022).
- 20. Kate Hardey, Robert Kantrowitz & Micah Desaire (Kirkland & Ellis), "<u>It's Time To Prescribe Frameworks For</u> <u>AI-Driven Health Care</u>," *Law360*, Oct. 26, 2023.
- 21. Darrell M. West, "Using AI and Machine Learning to Reduce Government Fraud," Brookings, Sept. 10, 2021.
- 22. Nick Peterson & Spencer Brooks, "<u>AI Use May Trigger False Claims Act's Public Disclosure Bar</u>," *Law360*, Nov. 8, 2023.
- Thomas C. King et al., "<u>Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions</u>," 26 Sci. & Eng'g Ethics 89 (2020).
- 24. Kevin Beasley, "Unlocking The Power of Predictive Analytics With AI," Forbes, Aug. 11, 2021.
- 25. See Beasley, "Unlocking The Power of Predictive Analytics With AI," supra.

Artificial Intelligence and Insurance-Part II, The Story Unfolds

- 26. Rashida Richardson, "<u>Addressing the Harmful Effects of Predictive Analytics Technologies</u>," *in #Tech2021: Ideas for Digital Democracy* (German Marshall Fund, Nov. 19, 2020).
- 27. Florian Möslein, "Robots in the Boardroom: Artificial Intelligence and Corporate Law," *in Research Handbook on the Law of Artificial Intelligence* (Woodrow Barfield & Ugo Pagallo eds., 2018).
- 28. This conclusion was written entirely by the generative AI platform TextCortex (https://textcortex.com/), based on the chatbot's review of this article.

Authors



Published by the American Bar Association ©2024. Reproduced with permission. All rights reserved. This information or any portion thereof may not be copied or disseminated in any form or by any means or stored in an electronic database or retrieval system without the express written consent of the American Bar Association.

ABA American Bar Association

https://www.americanbar.org/groups/litigation/resources/newsletters/insurance-coverage/artificial-intelligence-and-insurance-part-two/?login

LAWYERING IN THE AGE OF ARTIFICIAL INTELLIGENCE

Jonathan H. Choi,* Amy B. Monahan** & 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 using AI to complete legal tasks and correctly guessed the tasks for which GPT-4 were 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 affirmatively embrace AI tools and plan for a future in which they will become widespread.

^{*} 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, and the Harvard Law School Legal Profession Seminar. This research was supported by the Kommerstad Faculty Imagination Fund at the University of Minnesota Law School.

LAWYERING IN THE AGE OF AI

Contents

Introduction		3
I. Back	sground	8
II. Methodology		17
III. Results		21
IV. Limitations, Identification Assumptions, and Robustness Checks36		
V. Impl	lications	
A. In	nplications for the Future of Legal Services	
B. No	ormative Implications	46
1.	Lawyers and Law Firms	46
2.	Legal Clients	48
3.	Judges	50
4.	Law Schools and Law Students	51
Conclusion		55
Appendix		56
A. Tr	raining Materials	56
B. As	ssignments	58
C. Gr	raphs of Differences in Means	60
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 latter application is

¹ JOHN G. ROBERTS, 2023 YEAR-END REPORT ON THE FEDERAL JUDICIARY 5 (2023) (urging "caution and humility" among both judges and lawyers in the use of generative artificial intelligence); Roger E. Barton, How Will Leveraging AI Change the Future of Legal Services?, REUTERS (Aug. 23.2023.9:06AM), https://www.reuters.com/legal/legalindustry/how-will-leveraging-aichange-future-legal-services-2023-08-23; Daniel Farrar, To Future-Proof Their Firms, Attorneys Must Embrace AI, FORBES (July 13, 2023, 9:00 AM), https://www.forbes.com/sites/forbesbusinesscouncil/2023/07/13/tofuture-proof-their-firms-attorneys-must-embrace-ai/?sh=6282438b245b; Steve Lohr, A.I. is Coming for Lawyers, Again, N.Y. TIMES (April 10, 2023), https://www.nytimes.com/2023/04/10/technology/ai-is-coming-forlawyers-again.html; John Villasenor, How AI Will Revolutionize the BROOKINGS Practice of Law, INST. (March 20,2023),https://www.brookings.edu/articles/how-ai-will-revolutionize-thepractice-of-law.

² See, e.g., Jonathan H. Choi, Kristin E. Hickman, Amy B. Monahan, & Daniel Schwarcz, *ChatGPT Goes to Law School*, 71 J. LEGAL ED. 387, 388-89 (2022) (finding that exams 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 3-4 (Apr. 5, 2023) (unpublished manuscript) (on file with authors) (finding that, with limited prompting, GPT-4 passed the Uniform Bar Examination and significantly outperformed most human test-takers); Matthew Dahl, Varun Magesh, Mirac Suzgun, & Daniel E. Ho, Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models (2024) (working paper) (on file with authors) (concluding that "legal hallucinations are alarmingly prevalent" in analysis conducted by LLMs, but failing to account for the

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,

³ See, e.g., 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 ("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, FLA. BAR NEWS (March 2023), 13.https://www.floridabar.org/the-florida-bar-news/lawyers-and-ai-howlawyers-use-of-artificial-intelligence-could-implicate-the-rules-ofprofessional-conduct ("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); W. Bradley Wendel, The Promise and Limitations of Artificial Intelligence in

the Practice of Law, 72 OKLA. L. REV. 21, 24-26 (2019). 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.

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 HOWARTH, SHAPING THE BAR: THE FUTURE OF ATTORNEY LICENSING (2022) (arguing for significant reforms in the bar exam

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

many studies to date have suffered from methodological limitations, like non-blind grading of results⁷ 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. 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 currently available, GPT-4.⁹ 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 young attorneys.¹⁰ Prior to completing these tasks, study participants received several hours of training on how to use GPT-4 effectively, which we

because it has historically failed to test the skills that new lawyers need to represent clients while unfairly harming traditionally marginalized groups).

⁷ See Katz, Bommarito, Gao, & Arredondo, *supra* note 2, at 7 (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 also* Eric Martínez, Re-Evaluating GPT-4's Bar Exam Performance (June 12, 2023) (unpublished manuscript) (on file with authors) (critiquing OpenAI's claims that GPT-4 performed at the 90th percentile on the Uniform Bar Examination).

⁸ See Jonathan H. Choi & Daniel Schwarcz, AI Assistance in Legal Analysis, 72 J. LEGAL ED. (forthcoming 2024) [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 See OPENAI, GPT-4 TECHNICAL REPORT (2023), at https://arxiv.org/abs/2303.08774 (reporting that GPT-4's performance on various benchmarks exceeds the performance of prior generative AI models).

¹⁰ 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, 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).

patterned on real attorney training materials.¹¹ After participants completed the four assigned tasks, we blind-graded the results and tracked how long they took on each task.

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). 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%, -11.8%). 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). On the other hand, AI assistance reduced the amount of time that participants took to complete the tasks roughly uniformly regardless of their baseline speed.

We also surveyed participants on their perceptions of how access to GPT-4 impacted their work on the assigned legal tasks. We found that (again for the tasks on which GPT-4 was most useful) participants reported increased satisfaction from using it. Although they completed the survey before 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. 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 can selectively use AI in situations where it provides the greatest benefits.

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. 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.¹² Second, study participants only received several

¹¹ See infra Part II (describing the process for training study participants to use GPT-4 effectively to complete basic legal writing tasks).

¹² See infra Part IV. For instance, LexisNexis just recently launched an AI legal assistant that is built into its general-purpose search engine. See LEXISNEXIS, Lexis+AI, Transform Your Legal Work,

hours of training on GPT-4 before completing assigned tasks, whereas lawyers that use AI-based tools will continually refine their ability to skillfully use AI over the course of months or years.¹³ 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.¹⁴

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. 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. 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. Law schools should reassess when and how law students are trained to use AI, and when and how access to that tool is limited.

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 Three, we highlight and discuss our key results, which demonstrate that generative AI can

https://www.lexisnexis.com/en-us/products/lexis-plus-ai.page (last visited Jan. 9, 2024). Similarly, Thomson Reuters, the owner of Westlaw, recently acquired the firm Casetext in large part due to it 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-legaltech-provider-casetext-650-mln-2023-06-27. Westlaw \mathbf{is} currently working to integrate at least some of these capabilities into its Westlaw Precision product. See Press Release, Thomson Reuters Unveils Generative AI Strategy Designed to Transform the Future of Professionals (Nov. 1, 2023), https://www.thomsonreuters.com/en/pressreleases/2023/november/thomson-reuters-unveils-generative-aistrategy-designed-to-transform-the-future-of-professionals.html.

¹³ See infra Part III.

¹⁴ See, e.g., MCKINSEY & CO., WHAT'S THE FUTURE OF GENERATIVE AI? AN EARLY VIEW IN 15 CHARTS (2023), https://www.mckinsey.com/featured-insights/mckinseyexplainers/whats-the-future-of-generative-ai-an-early-view-in-15-charts (illustrating the pace of innovation in generative AI)

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 Four, 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 decisionmaking. 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 technology.¹⁵ 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.¹⁶ 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.¹⁷

Historically, these major legal tech innovations have improved lawyer efficiency rather than fundamentally altering the core skills needed to be an effective lawyer.¹⁸ For example, a lawyer with access to

¹⁷ 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).

¹⁸ Mark Fenwick et al., Legal Education in the Blockchain Revolution, 20 VAND. J. ENT. & TECH. L. 351, 357 (2017); Cass Sunstein, 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).

¹⁵ William G. Harrington, A Brief History of Computer-Assisted Legal Research, 77 LAW LIBR. J. 543, 553 (1985); 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); James A. Sprowl, Computer-Assisted Legal Research: Westlaw and Lexis, 62. AMER. BAR ASSOC. J. 320 (1976).

¹⁶ 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).

an easily searchable legal database can complete legal research in much less time than would be possible if they needed to search through hard copy indices. But the skill involved in analyzing and applying cases and statutes remains fundamentally the same.¹⁹ Similarly, e-discovery tools allow lawyers to automate the search function in discovery²⁰ but cannot provide the knowledge necessary to identify what must be produced and what is protected by privilege.

Even before the recent wave of progress in generative AI tools like ChatGPT, the rise of AI in legal tech was disrupting this historical pattern. For example, AI tools like predictive coding in e-discovery systems have become increasingly prominent in recent years. These tools allow a lawyer to code a sample of discovery documents, which are then used by an algorithm to identify other relevant documents.²¹ To a certain degree, tools such as these actually displace an attorney's work.²²

With each new innovation, lawyers have typically fretted about the implications for the legal profession and lawyer jobs.²³ If technology

²¹ See id. 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 a large portion of many lawyers' jobs involves the prediction of legal outcomes, which can be more accurately conducted with the aid of technology that leverages data about similar legal questions).

²² See Maura R. Grossman & Gordon V. Cormack, Quantifying Success: Using Data Science to Measure the Accuracy of Technology-Assisted Review in Electronic Discovery, in DATA DRIVEN LAW: DATA ANALYTICS AND THE NEW LEGAL SERVICES 127, 150-51 (Ed Walters ed., 2019) (finding that these "technology-assisted review" systems in ediscovery provided "significantly superior precision" compared to manual review). But see Emily S. Taylor Poppe, The Future Is Bright Complicated: AI, Apps & Access to Justice, 72 OKLA. L. REV. 185, 189 (2019) (arguing that displacement concerns are less significant when it comes to tasks that were already subject to outsourcing).

²³ See, e.g., RICHARD SUSSKIND & DANIEL SUSSKIND, THE FUTURE OF THE PROFESSIONS: HOW TECHNOLOGY WILL TRANSFORM THE WORK OF HUMAN EXPERTS 66 (2015); Daniel Martin Katz, *Quantitative Legal*

¹⁹ See Raymond H. Brescia et al., Embracing Disruption: How Technological Change in the Delivery of Legal Services Can Improve Access to Justice, 78 ALB. L. REV. 553, 568 (2014).

²⁰ See John O. McGinnis & Russell G. Pearce, *The Great Disruption: How Machine Intelligence Will Transform the Role of Lawyers in the Delivery of Legal Services*, 82 FORDHAM L. REV. 3041, 3047-48 (2014)

allowed the same work to be done in less time,²⁴ or could replace lawyers altogether for certain tasks,²⁵ 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.²⁶ 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.²⁷ Moreover, it is possible that task automation could also increase the demand for lawyers, either because the lower cost of legal services increases the overall quantity of legal services provided (induced

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, 909 (2013); McGinnis & Pearce, supra note 20, at 3047-48; Dana A. Remus & Frank Levy, Can Robots Be Lawyers: Computers, Lawyers, and the Practice of Law, 30 GEO J. LEGAL ETHICS 501, 501 (2017); Tanina Rostain, Robots versus Lawyers: A User-Centered Approach, 30 GEO J. LEGAL ETHICS 559, 560 (2017); Sean Semmler & Zeeve Rose, Artificial Intelligence: Application Today and Implications Tomorrow, 16 DUKE L. & TECH. REV. 85, 86 (2017); Harry Surden, Machine Learning and Law, 89 WASH. L. REV. 87, 87 (2014); David C. Vladeck, Machines Without Principals: Liability Rules and Artificial Intelligence, 89 WASH. L. REV. 117 (2014); John Markoff, Armies of Expensive Lawyers, Replaced by Cheaper Software, N.Y. TIMES (Mar. 4, 2011), https://www.nytimes.com/2011/03/05/science/05legal.html; JAMES E. MOLITERNO, THE AMERICAN LEGAL PROFESSION IN CRISIS: RESISTANCE AND RESPONSES TO CHANGE 208 (2013).

²⁴ See Mark Fenwick et al., Legal Education in the Blockchain Revolution, 20 VAND. J. ENT. & TECH. L. 351, 357 (2017).

²⁵ Christopher A. Suarez, *Disruptive Legal Technology, COVID-*19, and Resilience in the Profession, 72 S.C. L. REV. 393, 404 (2020).

²⁶ See Gillian Hadfield, Legal Barriers to Innovation: The Growing Economic Cost of Professional Control Over Corporate Legal Markets, 60 STAN. L. REV. 1689, 1724-25 (2008).

²⁷ See, e.g., SUSSKIND & SUSSKIND, supra note 23, at 66-67; McGinnis & Pearce, supra note 20, at 3047-48; Raymond H. Brescia et al., Embracing Disruption: How Technological Change in the Delivery of Legal Services Can Improve Access to Justice, 78 ALBANY L. REV. 553, 553 (2015); Elinor R. Jordan, Point, Click, Green Card: Can Technology Close the Gap in Immigrant Access to Justice?, 31 GEO. IMMIGR. L.J. 287 (2017); Elliott Vinson & Samantha A. Moppett, Digital Pro Bono: Leveraging Technology to Provide Access to Justice, 92 ST. JOHNS L. REV. 551 (2018); J.J. Prescott, Improving Access to Justice in State Courts with Platform Technology, 70 VAND. L. REV. 1993, 1994 (2017).

10

demand) or because automation creates new tasks for which human labor is an important complement (what Daron Acemoglu and Pascal Restrepo have called a "reinstatement effect").²⁸

Similar dynamics exist in recent discussions of how increasingly capable LLMs like GPT-4²⁹ will impact the legal profession. At the same time, LLMs like GPT-4 seems to represent a qualitatively different type of technological advance from those that came before. As a result, many have speculated that these LLMs will lead to true revolution in the practice of law,³⁰ radically changing market demand for human lawyers.³¹

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

²⁹ See, e.g., Erin Mulvaney & Laura Webber, End of the Billable Hour? Law Firms Get on Board with Artificial Intelligence, WALL ST. J. (May 11, 2023, 11:00 AM), https://www.wsj.com/articles/end-of-thebillable-hour-law-firms-get-on-board-with-artificial-intelligence-17ebd3f8.

³¹ 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 and GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models (Aug. 22, 2023) (unpublished manuscript) (on file with authors) (evaluating the potential labor market effects of LLMs like GPT-4).

11

²⁸ Daron Acemoglu & Pascal Restrepo, Automation and New Tasks: How Technology Displaces and Reinstates Labor, 33 J. ECON. PERSPECTIVES 3, 3 (2019) (describing the countervailing actions of the "displacement effect," where labor is replaced by automation, and the reinstatement effect). See also Daron Acemoglu et al., Artificial Intelligence and Jobs: Evidence from Online Vacancies, 40 J. LABOR ECON. 293 (2022) (analyzing the effect of AI on jobs but finding them too small for firm conclusions); Daron Acemoglu, The Simple Macroeconomics of AI (2024) (working paper) (on file with authors) (analyzing the macroeconomic effects of recent AI developments).

³⁰ 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 (2018) (speculating that AI will substantially transform the work of lawyers in the future).

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,³² bar examinations,³³ and in answering discrete legal questions.³⁴ Other non-empirical research considers the ethical implications of using such technology in the practice of law,³⁵ how artificial intelligence may change the skills needed to be a successful lawyer,³⁶ and how law firms may begin to compete on the basis of technological expertise.³⁷

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 90th percentile on the Uniform Bar Examination³⁸

³⁵ See, e.g., Katherine Medianik, Artificially Intelligent Lawyers: Updating the Model Rules of Professional Conduct in Accordance with the New Technological Era, 39 CARDOZO L. REV. 1497 (2018); Brian L. Frye, Should Using an AI Text Generator to Produce Academic Writing Be Plagiarism?, 33 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 946, 958 (2023).

³⁶ 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); Suarez, supra note 25, at 396.

³⁷ Bruce A. Green & Carole Silver, *Technocapital@biglaw.com*, 18 NW. J. TECH. & INTELL. PROP. 265, 282-308 (2021).

³⁸ Katz, Bommarito, Gao, & Arredondo, *supra* note 2, at 7. This result extended both to the multiple-choice portion of the exam as well as to the open-ended essay components of the exam. *Id.* at 2. Although the authors did not use any prompt-engineering strategies to generate

³² Choi, Hickman, Monahan, & Schwarcz, *supra* note 2 (testing the performance of GPT-3.5 alone on law school exams). *See also* Andrew Blair-Stanek et al., GPT-4's Law School Grades: Con Law C, Crim C-, Law & Econ C, Partnership Tax B, Property B-, Tax B (May 24, 2023) (unpublished manuscript) (on file with authors) [hereinafter *GPT-4's Law School Grades*]; Margaret Ryznar, *Exams in the Time of ChatGPT*, 80 WASH. & LEE L. REV. ONLINE 305 (2023) (reporting mixed results).

³³ Katz, Bommarito, Gao, & Arredondo, *supra* note 2, at 7.

³⁴ John Ney et al., Large Language Models as Tax Attorneys: A Case Study in Legal Capabilities Emergence, 381 PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A: MATHEMATICAL, PHYSICAL AND ENGINEERING SCIENCES (forthcoming 2023); Andrew Blair-Stanek, Nils Holzenberger, & Benjamin Van Durme, OpenAI Cribbed Our Tax Example, But Can GPT-4 Really Do Tax?, 180 TAX NOTES FED. 1101, 1105 (2023) [hereinafter OpenAI Cribbed Our Tax Example].

(although scholars have subsequently raised methodological doubts about this claim³⁹). In another study evaluating AI-generated answers to law school exam questions, researchers found that although exams drafted by GPT-3.5 often included solid explanations of basic legal rules and strong organization and composition, they also often struggled to identify relevant issues and tended to only superficially apply rules to facts as compared to real law students.⁴⁰ Perhaps most interestingly, a later study examining GPT-4 assistance on law school exams, where

³⁹ Martínez, *supra* note 7 (discussing potential methodological issues with the initial finding that GPT-4 surpassed the bar exam score of 90% of human test takers). In addition, the authors in the Katz et al. study did not grade GPT-4's performance blind and did not have experience grading bar exams, raising concerns about subjective bias in evaluation. *Id*.

⁴⁰ Choi, Hickman, Monahan, & Schwarcz, *supra* note 2, at 388-89 (testing the performance of GPT-3.5 alone on law school exams). *See also* Blair-Stanek et al., *supra* note 32; Ryznar, supra note 32, at 305 (reporting mixed results).

In other disciplines, GPT has been found to be a proficient and sometimes superior test taker as compared to humans. See Harsha Nori et al., Capabilities of GPT-4 on Medical Challenge Problems (Apr. 12, 2023) (unpublished manuscript) (on file with authors) (finding that GPT-4, without any specialized prompting passes a range of medical exams and out-performs both ChatGPT and LLM models specifically fine-tuned on medical knowledge); John C. Lin et al., Comparison of GPT-3.5, GPT-4. and Human User Performance on a Practice Ophthalmology Written Examination, 37 NATURE: EYE 3694, 3694 (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, 1353 (2023) (finding that both GPT-4 and GPT-3.5 pass neurosurgery practice board exams at rates comparable to neurosurgery residents); Hanmeng Liu et al., Evaluating the Logical Reasoning Ability of ChatGPT and GPT-4 (May 5, 2023) (unpublished manuscript) (on file with authors); Vinay Pursnani, Yusuf Sermet & Ibrahim Demir, Performance of ChatGPT on the US Fundamentals of Engineering Exam: Comprehensive Assessment of Proficiency and Potential Implications for Professional Environmental Engineering Practice (Apr. 20, 2023) (unpublished manuscript) (on file with authors).

13

multiple choice answers, they slightly modified essay questions by presenting each sub-question in an independent prompt and by "lightly correcting the language" in the prompt so that it formed a complete sentence. *Id.* at 7.

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.⁴¹

Outside of the exam context, little evidence exists on how access to LLM tools like GPT-4 might impact lawyers' or law students' abilities to complete legal tasks. Tax scholars have tested GPT-4's ability to answer questions about federal tax law, generally finding low accuracy with basic prompting (roughly 30% in two separate studies) to 70%-90% accuracy with significant human assistance (particularly prompting with hand-selected correct sources).⁴² Many scholars have anecdotally tested GPT's capabilities, including a series of YouTube videos that illustrate GPT-4's capabilities in various legal contexts.⁴³ 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.⁴⁴

In areas other than law, we see the same general focus on exam performance rather than studies of realistic tasks. And as with law, the exam results are mixed. Whereas exams generated by ChatGPT were rated as "outstanding" in economics⁴⁵, they achieved more middling results in computer programming and medical education,⁴⁶ and

⁴¹ Choi & Schwarcz, AI Assistance in Legal Analysis, supra note 8.

 $^{^{42}}$ Ney et al., supra note 34, at 381; Blair-Stanek et al., supra note 34, at 1105.

⁴³ *Harry Surden*, YOUTUBE (Jan. 8, 2024), https://www.youtube.com/@harrysurden3116.

⁴⁴ Harry Surden, *ChatGPT Analyzes Copyright Law*, YOUTUBE (March 22, 2023), https://www.youtube.com/watch?v=nqZcrhR8yPU.

⁴⁵ Wayne Geerling et al., *ChatGPT Has Aced the Test of Understanding in College Economics: Now What?*, 68 AMER. ECON. 233, 233 (2023) (finding that GPT ranked in the 91st percentile for Microeconomics and the 99th percentile for Macroeconomics when compared to college students taking the Test of Understanding in College Economics).

⁴⁶ Tiffany H. Kung et al., *Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models*, PLOS DIGITAL HEALTH (2023) (reporting that ChatGPT performed "at or near the passing threshold" on the United States Medical Licensing Exam). *See also* Peter Lee, Sebastien Bubeck, & Joseph Petro, *Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine*, 388 NEW ENG. J. MED. 1233, 1233 (2023).

"unsatisfactory" results in fields like mathematics and psychology.⁴⁷ Common problems with ChatGPT-drafted exams included inaccurate, unreliable, and outdated information.⁴⁸ These studies vary significantly in the methods they use to test LLM performance. Some test the performance of AI acting alone, where a question or prompt is entered into an LLM and its answer is evaluated without modification. Other studies examine the value of AI *assistance*, where a human subject uses an LLM on various tasks or subtasks and then reviews, edits, or otherwise refines those results to produce a final work product.

Outside of the exam setting, a small number of studies have evaluated how AI can improve human performance at non-legal professional writing tasks.⁴⁹ One study found that giving collegeeducated 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.⁵⁰ On the other hand, other empirical work

 48 See Lo, supra note 47, at 5-6.

⁴⁹ There are some recent papers that evaluate how access to generative AI can improve professionals' ability to perform non-writing tasks, like computer coding. *See* Sida Peng et al., The Impact of AI on Developer Productivity: Evidence from GitHub Copilot (Feb. 13, 2023) (unpublished manuscript) (on file with authors). None of these studies evaluate how more sophisticated prompting techniques can impact results.

⁵⁰ Shakked Noy & Whitney Zhang, *Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence*, 381 SCIENCE 187, 187 (2023). To reach this conclusion, the experimenters recruited over 400 participants in five professional categories: grant writers, consultants, data analysts, human resource professionals, and managers. Participants were then tasked with completing two short writing assignments comparable to those they would complete in their professional settings, such as drafting press releases, short reports or emails. After completing the first writing assignment, half of the participants were given access to ChatGPT for the second writing assignment. The study found that participants who were provided with access to ChatGPT completed their writing tasks faster and produced higher quality work than participants who were not provided access to

⁴⁷ See Chung Kwan Lo, What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature ,13 EDUC. SCI. 410, 410 (2023); see also Lakshmi Varanasi, ChatGPT Could Be a Stanford Medical Student, a Lawyer, or a Financial Analyst. Here's a List of Advanced Exams the AI Bot Has Passed So Far, BUSINESS INSIDER (Nov 5, 2023, 4:47 PM) https://www.businessinsider.com/list-here-are-theexams-chatgpt-has-passed-so-far-2023-1.

has suggested that human use of AI to assist with certain tasks can undermine humans' incentives to take care.⁵¹

One of the most extensive studies of AI-assistance in knowledgeintensive work examined the effect of AI-assistance on a range of work tasks common within the field of high-level management consulting.⁵² The results show that AI is remarkably capable of increasing both quality and productivity on certain types of tasks but not others, even where the tasks are considered of similar difficulty. Specifically, consultants completing a series of tasks that involved conceptualizing and developing new product ideas significantly improved both the quality and speed of their work with the assistance of AI.⁵³ 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.⁵⁴ 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.⁵⁵ The study also found, consistent with studies conducted by Choi & Schwarcz and Noy & Zhang, that the most significant

⁵⁵ *Id.* at 10, 15.

this tool. Moreover, the participants who performed relatively poorly on the initial task (which took place prior to being instructed how to use ChatGPT) disproportionately benefited from access to AI, receiving both higher quality scores and taking decreased amounts of time to complete their writing task. By contrast, access to ChatGPT did not improve the quality of work for participants who scored well in the initial writing task, though it did increase the speed at which they could produce that work.

⁵¹ Fabrizio Dell'Acqua, Falling Asleep at the Wheel: Human/AI Collaboration in a Field Experiment on HR 1 (Dec. 2, 2021) (unpublished manuscript) (on file with authors) [hereinafter *Falling Asleep at the Wheel*].

⁵² 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*].

⁵³ *Id.* at 9-10. *See also* Karan Girotra et al., Ideas are Dimes a Dozen: Large Language Models for Idea Generation in Innovation (July 10, 2023), (unpublished manuscript) (on file with authors) (finding that GPT-4 can generate ideas faster and cheaper than college students at an elite university).

⁵⁴ Dell'Acqua et al., *Navigating the Jagged Technological Frontier*, *supra* note 52, at 13-15.

beneficiaries of AI assistance were lower-skilled participants.⁵⁶ However, in contrast to Choi & Schwarcz, the study found performance improvements even among those in the top half of skill rankings.⁵⁷ 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.⁵⁸ Researchers also found that participants who blindly adopted AI outputs suffered a decrease in performance compared to those not using AI assistance at all.⁵⁹

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

Our study aims to move the literature forward by evaluating the effect of GPT-4 assistance, in terms of both quality and efficiency, on four 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 2023.⁶⁰

 $^{^{56}}$ Id. at 11 (finding a 43% increase in performance among those ranked in the bottom half of skill level).

 $^{^{57}}$ Id. at 11 (finding a 17% increase in performance among those ranked in the top half).

⁵⁸ *Id.* at 12. *See also* Leonard Boussioux et al., The Crowdless Future? How Generative AI is Shaping the Future of Human Crowdsourcing (Aug. 8, 2023), (unpublished manuscript) (on file with authors) (similarly finding that GPT-4 may decrease some forms of creativity and novelty compared to purely human outputs).

⁵⁹ Dell'Acqua et al., *Navigating the Jagged Technological Frontier*, *supra* note 52, at 17.

⁶⁰ The University of Minnesota Law School is one of the top law schools in the country, currently ranked 16th in the U.S. News ranking of law schools. 2023 Best Law Schools, U.S. NEWS, https://www.usnews.com/best-graduate-schools/top-law-schools/law-rankings (last visited Aug. 5, 2023).

Well over 100 students expressed interest in participating in the study.⁶¹ We initially enrolled the first sixty such volunteers and placed the remaining volunteers on a waitlist.⁶² Over the duration of the study, 22 of the participants dropped out because they were unable to complete the entirety of the experiment; as they did so, we replaced them with new participants from the waitlist to ensure that we achieved roughly the target number of sixty study participants. Ultimately 59 students completed the experiment.

During the enrollment process, we gathered basic information about study participants, including their first-semester first-year law school GPA and their anticipated graduation year.⁶³ 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.⁶⁴ Doing so required

⁶¹ One of the co-authors sent a recruiting email to the entire University of Minnesota Law School student body in April 2023. The email explained that we were recruiting "current JD students, including class of 2023 graduates, for participation in a study that examines the use of artificial intelligence tools, specifically GPT- 4, to assist with basic lawyering tasks." To participate in the study, students or graduates would need to be available to work for up to 15 hours total during June 2023. The email also noted that the work could be completed remotely and on participants' own time-schedules and that participants who completed the study receive \$300 in compensation for their time.

⁶² This experimental design was approved by the University of Minnesota's IRB. Participants agreed to participate after reviewing and agreeing to an IRB-approved consent form.

⁶³ We also collected contact information, including email and mailing address, 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.

⁶⁴ This training drew heavily on previous work by two of us. See 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,

students to watch approximately two hours of training videos and to complete several short exercises using GPT-4. The training included both general techniques on how to prompt GPT-4 effectively (for example, by breaking down legal analysis into pieces and supplying relevant legal rules or sources) and how to use it specifically in litigation and transactional settings. It focused on how to apply active lawyering skills while using AI, rather than mechanically relying on the output of GPT-4. For example, we instructed participants to first assess assignments on their own before using GPT-4 to generate answers. 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

presentations for in-house legal teams, and presentations for lawyers working at large law firms.

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 prevent participants from spending more time than necessary on a task in order to maximize their compensation. Participants also received the following instructions for each task:

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

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

To provide access to GPT-4 to participants, we created a central ChatGPT "clone" website using the GPT-4 API, and gave students access to that website.⁶⁵ This clone website had a nearly identical user interface and used the same system prompt as the real ChatGPT Plus with GPT-4.

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

At the completion of the experiment, all participants were asked to take an anonymous survey regarding their experience. Although the survey was anonymous on a per-respondent basis, we tracked responses separately for Groups A and B, allowing us to register how each group felt on average about their respective assignments. We pre-registered our methods and hypotheses prior to analyzing our results; the preregistration statement is archived with the Open Science Foundation.⁶⁶

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.

⁶⁵ 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.

⁶⁶ 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/5yzj3. 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).

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.⁶⁷ 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 generated was approximately two thirds of the difference between a B and a B+. The results also show substantial variation between tasks. On the client memo and EE handbook task, respondents saw, on average, a near zero effect on performance from using GPT-4.

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

	No GPT-4	With GPT-4	Difference	<i>p</i> -value
	(Std. Dev.)	(Std. Dev.)	(95% CI)	
Complaint	3.14	3.31	0.17	0.0862
Drafting	(0.59)	(0.50)	(-0.03, 0.37)	
Contract	3.00	3.24	0.24	0.0060
Drafting	(0.56)	(0.40)	(0.07, 0.41)	
EE	3.20	3.26	0.07	0.3532
Handbook	(0.41)	(0.39)	(-0.07, 0.21)	
Client	2.92	2.85	-0.07	0.5980
Memo	(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).⁶⁸ Figure 17 through Figure 20 in the Appendix show the bootstraps for the difference in means for groups with and without access to GPT, showing that only contract drafting showed a statistically significant increase in performance at the 95% level.

⁶⁷ All confidence intervals in this Article were generated using empirical bootstraps with 10,000 iterations.

⁶⁸ All figures in this Article were generated using the SciPy package in Python. Density plots were generated using Gaussian Kernel Density Estimation using the gaussian_kde package in SciPy, applying Scott's rule of thumb to determine bandwidth. See Adriano Z. Zambom & Ronaldo Dias, *A Review of Kernel Density Estimation with Applications to Econometrics*, 5 INT'L ECON. REV. 20, 29-31 (2013).



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 (again in percentage terms) occurs in the task for which GPT-4 was the least useful (client memo).

(minutes)					
	No GPT-4	With GPT-4	Difference	%	р-
	(Std. Dev.)	(Std. Dev.)	(95% CI)	Difference	value
Complaint	160.69	122.00	-38.77	24.1%	0.0018
Drafting	(72.38)	(66.80)	(-64.00, -13.36)		
Contract	69.72	47.59	-22.40	32.1%	0.0000
Drafting	(32.00)	(31.09)	(-33.71, -10.91)		
EE 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)		

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

Figure 5 through Figure 8 below show the distributions of the amount of time that participants took on each task. Figure 21 through Figure 24 in the Appendix show bootstraps for the differences in means between groups, showing that the decrease in the time participants took on every task is statistically significant at the 95% level.



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

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





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





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 45-degree angle where x = y. The graphs are separated based on which task was used as Task 2.

We can then take the two tasks that each participant completed without access to AI and use them to graph another line, showing how their performance on a task without GPT-4 (on the *x*-axis) predicts performance with access to GPT-4 (on the *y*-axis). This is the red line in the figures below.⁶⁹ 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.

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



2.0

Grade (Task 1)

1.5

3.0

2.5

3.5

4.0

0.5

 $\begin{array}{c} 0.0\\ 0.0\end{array}$

0.5

1.0

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

In sum, where assistance from GPT-4 is beneficial at all, it seems to benefit the worst performers the most, providing little or no benefit to top performers. Table 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 95% level.

	No GPT	With GPT	Difference
	(95% CI)	(95% CI)	(95% CI)
Complaint	0.66	0.16	0.50
Drafting	(0.35, 0.95)	(0.00, 0.28)	(0.20, 0.84)
Contract	0.56	0.19	0.37
Drafting	(0.33, 0.80)	(-0.06, 0.20)	(0.22, 0.74)
Employee	0.01	0.06	-0.05
Handbook	(-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)

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

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 95% level. Our results therefore suggest that GPT-4 has the potential to reduce time spent on task for lawyers of all ability levels.



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







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





Table 4 reflects these results.

Tuble 1. Stope of GIT T on Terrormance (Grauc)			
	No GPT	With GPT	Difference
	(95% CI)	(95% CI)	(95% CI)
Complaint	0.63	0.60	0.03
Drafting	(0.39, 0.90)	(0.26, 0.88)	(-0.32, 0.48)
Contract	0.74	0.40	0.34
Drafting	(0.52, 0.96)	(0.12, 0.75)	(-0.08, 0.68)
Employee	0.28	0.34	-0.06
Handbook	(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?

LAWYERING IN THE AGE OF AI

- 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 26: Survey Results by Question



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 60 participants each completing 4 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.⁷⁰ 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.

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 2 hours of training prior to task completion. Scaling the sample size as well 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

⁷⁰ Noy & Zhang, *supra* note 50, at 187.

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 2 tasks with AI assistance and 2 tasks without. (We promised this to participants to give them the opportunity to contrast 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 1L Fall GPA. 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 1L Fall GPA and do not suggest any statistically significant differences between the two groups.

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

Table 5: Group A and Group B Individual Characteristics

In addition, we conducted Kolgorov-Smirnov tests to estimate the likelihood that the class years for Group A and Group B, and the 1L Fall GPAs for Group A and Group B, were drawn from the same distribution. The Kolgorov-Smirnov statistic for class year was 0.14 (p = 0.95) and for 1L Fall GPA was 0.10 (p = 1.00), again 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),⁷¹ 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. 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

⁷¹ See Donald B. Rubin, *Which Ifs Have Causal Answers*, 81 J. AM. STAT. ASSOC. 961, 961 (1986) (providing a technical mathematical definition of SUTVA).
that they accurately capture 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 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 a broad array of legal tasks without adversely affecting (or even slightly improving) the quality of that work product. Moreover, they suggest that young lawyers provided with access to AI to facilitate their work accurately appreciate these benefits of AI, find that access to AI tends to enhance their work satisfaction, and generally become more enthusiastic about using AI to facilitate their work as they gain experience doing so.

Standing alone, these results suggest that generative AI will almost certainly become a vital tool for many lawyers in the near future, comparable to more familiar legal-tech tools like Westlaw, Lexis and ediscovery software.⁷² Indeed, this trend has already begun, with some lawyers and law firms proactively embracing generative AI.⁷³ 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.⁷⁴ Currently available law-

40

⁷² See supra Part I (describing the evolution of legal technology tools, including searchable online databases like Westlaw and Lexis).

⁷³ 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-4078-9b3eddfce710a6e2; Emily Hinkley, Mishcon de Reya Is Hiring an 'Engineer' to Explore How Its Lawyers Can Use ChatGPT, LEGAL CHEEK (Feb. 16, 2023, 8:35:00 AM), https://www.legalcheek.com/2023/02/mishcon-dereya-is-hiring-an-engineer-to-explore-how-its-lawyers-can-use-chatgpt.

⁷⁴ For instance, the firm Casetext recently launched a product known as CoCounsel, which "does document review, legal research memos, deposition preparation, and contract analysis in minutes—with results you can trust." CASETEXT, *Meet Your New AI Legal Assistant*, https://casetext.com [https://perma.cc/5SDR-PG3S]. Within months of CoCounsel's launch, the legal tech giant Thomson Reuters purchased Casetext for \$650 Million. *See, e.g., Thomson Reuters to Acquire Legal AI Firm Casetext for \$650 Million*, REUTERS (June 27, 2023), https://www.reuters.com/markets/deals/thomson-reuters-acquire-legaltechprovider-casetext-650-mln-2023-06-27.

specific tools offer lawyers vastly superior capabilities than the generalpurpose 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.⁷⁵ RAG, the latter approach, allows generative AIs to retrieve relevant content from large legal databases and to use this material to inform its responses.⁷⁶ Combined, these two techniques substantially reduce hallucinations and improve the quality of AIgenerated output.⁷⁷

A second way in which our results understate the potential of AI to improve the efficiency of legal services is that our study participants had limited experience using this technology. In total, participants in our study received a couple hours of online training before attempting to use this technology to craft answers to two of the four assignments they completed while participating in the study.⁷⁸ Not surprisingly, participants did not believe that this training fully equipped them to use

⁷⁸ See supra Part II (describing training of study participants).

⁷⁵ 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-promptengineering; Tyler Cowen & Alexander T. Tabarrok, How to Learn and Teach Economics with Large Language Models, Including GPT (March 17, 2023) (unpublished manuscript) (on file with authors). For promptengineering advice that is specific to the legal setting, see Schwarcz & Choi, *AI Tools for Lawyers, supra* note 64.

⁷⁶ See generally Patrick Lewis et al., *Retrieval-Augmented* Generation for Knowledge-Intensive NLP Tasks, 33 ADVANCES IN NEURAL INFO. PROCESSING SYS. 9459 (2020). For discussion of how tools like Casetext use RAG, see CASETEXT, With AI You Get What You Give (Aug. 2, 2023), https://casetext.com/blog/prompt-engineering-best-aioutput ("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.").

⁷⁷ 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.

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.⁷⁹ 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.⁸⁰

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.⁸¹ 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 2022.⁸² 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 Summer of 2023.⁸³ For instance, due to model limitations, our participants were

⁸¹ See The Great Acceleration: CIO Perspectives On Generative AI, MIT TECH REV. (July 18, 2023), https://www.technologyreview.com/2023/07/18/1076423/the-greatacceleration-cio-perspectives-on-generative-ai.

⁸² See, e.g., Katz, Bommarito, Gao, & Arredondo, supra note 2, at 7; Choi & Schwarcz, AI Assistance in Legal Analysis, supra note 8; Blair-Stanek et al., supra note 32. For examples demonstrating that GPT-4 outperforms ChatGPT in other fields, see Chung Kwan, What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature, 13 EDUC. SCI. 410 (2023); David A. Wood et al., The ChatGPT Artificial Intelligence Chatbot: How Well Does It Answer Accounting Assessment Questions?, 2023 ISSUES IN ACCT. EDUC. 1; Harsha Nori et al., Capabilities of GPT-4 on Medical Challenge Problems (Apr. 12, 2023) (unpublished manuscript) (on file with authors); Alejandro Lopez-Lira & Yuehua Tang, Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models (May 12, 2023) (unpublished manuscript) (on file with authors).

⁸³ 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*

⁷⁹ See supra Part III (discussing survey results of study participants).

⁸⁰ See Schwarcz & Choi, AI Tools for Lawyers, supra note 64, at 5 ("The quickest route to proficiency with LLMs is the same route to Carnegie Hall: practice, practice, practice.").

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.⁸⁴ 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.⁸⁵ Given the relative homogeneity of our

at DevDay, OPENAI (Nov. 6, 2023), https://openai.com/blog/new-modelsand-developer-products-announced-at-devday.

⁸⁴ See Part III, supra.

⁸⁵ See Choi & Schwarcz, 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 "worstperforming students benefited enormously from AI, with gains of approximately 45 percentile points," while "the best-performing students received worse grades when given access to AI, experiencing declines of approximately 20 percentile points"). For literature outside of the legal setting finding uneven quality gains from access to AI based on the baseline skill of workers, see Noy & Zhang, supra note 45, at 187 (finding that giving college-educated professionals access to AI improved the performance of less skills workers more than high skilled workers); Erik Brynjolfsson, Danielle Li & Lindsey R. Raymond, Generative AI at Work, (Apr. 12, 2023) (unpublished manuscript) (on file with authors) (finding that giving customer service agents access to AI improved the capabilities of less skilled agents more than highly skilled agents). See also Dell'Acqua, Falling Asleep at the Wheel, supra note 51, at 1 (finding that giving professional recruiters access to high quality AI harmed

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.⁸⁶ It may also be the 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

humans' ability to assess job applications relative to giving them access to less high-quality AI tools).

⁸⁶ See Choi & Schwarcz, AI Assistance in Legal Analysis, supra note 8 (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).

four of the legal tasks we selected for the study necessarily shared certain features given our experimental design: they required a written work product, necessitating little if any independent research, that could be completed in between 1 to 5 hours of time, and that were reasonably appropriate for law students. These constraints, of course, do not apply to the immense range of tasks that real lawyers may need to complete. The features of some lawyer tasks—such as negotiating complex deal terms or crafting high-stakes legal briefs—almost certainly make them less amenable to assistance from AI. Meanwhile, many other legal tasks are likely to be much more dramatically impacted by the availability of AI than those that we focused on in our experimental setting. One important example involves the simple act of summarizing large and complex documents, such as deposition transcripts. General purpose AIs are particularly adept at summarizing complex and dense material, and specialized AI tools like CoCounsel use basic prompt engineering strategies to improve the reliability and verifiability of these efforts.⁸⁷ 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.

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,⁸⁸ a tool

analysis/X15S722S000000?bna_news_filter=bloomberg-law-

⁸⁷ See What It Takes To Build An AI Legal Assistant Lawyers Can Rely On, CASETEXT (May 12, 2023), https://casetext.com/blog/buildingan-ai-legal-assistant-lawyers-can-trust.

⁸⁸ 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#jcite (reporting that short-term job satisfaction was down while rates of burnout were up, particularly among junior and mid-level associates); AM. BAR ASSOC. NAT'L TASK FORCE ON LAWYER WELL-BEING, THE PATH TO LAWYER WELL-BEING: PRACTICAL RECOMMENDATIONS FOR POSITIVE CHANGE 7 (2017),

that has the potential to increase lawyer wellbeing, 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 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

https://www.americanbar.org/content/dam/aba/images/abanews/ThePat hToLawyerWellBeingReportRevFINAL.pdf (reporting low rates of wellbeing among early career lawyers and law students).

tools for lawyers, into its practice.⁸⁹ Numerous other large law firms have also embraced this tool in recent months, though many have been reluctant to publicly acknowledge this.⁹⁰ Other large global law firms including Allen & Overy—have incorporated a competing generative AI tool, Harvey, into their practice.⁹¹ Still other firms have taken a different approach, hiring their own AI experts to develop proprietary and firmspecific generative AIs that are not available to competitors.⁹²

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 lawyer who relied on ChatGPT to author a brief without double-checking the resulting output. The generative AI proceeded to hallucinate the existence of several cases, and then to insist on questioning from the lawyer that these cases were real. Not surprisingly, the unwitting lawyer was publicly excoriated by the judge in a hearing that was reported on widely by the media and that drew widespread attention from the bar.⁹³

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

⁹⁰ See Press Release, Top Global Law Firm DLA Piper Announces Addition of CoCounsel to Enhance Practice and Client Services (March 23, 2023), https://casetext.com/blog/law-firm-dla-piper-announcescasetext-cocounsel.

⁹¹ Charlotte Johnstone, *MacFarlanes Joins List of Firms Adopting Harvey AI*, ALM (Sept. 21, 2023, 4:28 AM), https://www.law.com/international-edition/2023/09/21/macfarlanes-joins-list-of-firms-adopting-harvey-ai.

⁹² 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, 7:00 AM), https://www.forbes.com/sites/lanceeliot/2023/10/17/prestigioussymposium-on-ai-lawyering-reveals-keen-insights-including-the-ardentdebate-on-whether-to-use-generative-ai-in-law-school-education.

⁹³ 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-lawsuitchatgpt.html.

⁸⁹ See Press Release, DLA Piper to Utilize CoCounsel, The Groundbreaking AI Legal Assistant Powered By OpenAI Technology (March 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.

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 made by AIs, (ii) experimenting with different prompting strategies when using general purpose AIs, including few-shot and grounded prompting, (iii) assessing legal issues and tasks independently of AI, and (iv) avoiding entering any confidential information into general purpose AIs that do not include trustworthy assurance of confidentiality.⁹⁴ 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.⁹⁵ 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.⁹⁶ Some lawyers may be inclined to accomplish this simply be resisting incorporating generative AI into

⁹⁴ See Schwarcz & Choi, AI Tools for Lawyers, supra note 64.

⁹⁵ 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).

⁹⁶ See Lisa G. Lerman, A Double Standard for Lawyer Dishonesty: Billing Fraud Versus Misappropriation, 34 HOFSTRA L. REV. 847, 848 (2006).

their workflows, citing some of the risks of this technology described above. Others may explain to clients that their use of generative AI has allowed them to invest their scarce time into other ways of protecting the 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.⁹⁷ The efficiencies associated with generative AI are virtually certain to shift the calculations associated with this make-buy decision. Most obviously, generative AI should allow clients to complete a larger percentage of routine legal work in house. Additionally, the uncertainty that generative AI introduces in how long legal work should take also counsels in favor of moving relatively routine work from external counsel to in house, as that shift should allow firms to better calibrate these expectations internally, where principal-agent problems are reduced.

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

⁹⁷ See John Armour & Mari Sako, AI- Enabled Business Models in Legal Services: From Traditional Law Firms to Next-Generation Law Companies, 7 J. PROFESSIONS & 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,⁹⁸ 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.⁹⁹ Other judges go further, requiring lawyers to specially certify the accuracy of any filings for which generative AI has been used.¹⁰⁰ And several judges have even prohibited lawyers that practice before them from using any generative AI to assist them with writing legal filings.¹⁰¹

In our view, our results suggest that such aggressive attempts to limit or complicate lawyers' use of generative AI are misguided.¹⁰² Generative AI has the capacity to allow lawyers to better serve their clients by producing work product more efficiently, thus reducing

⁹⁸ See, e.g., Ella Lee, Michael Cohen Gave Lawyer Fraudulent Case Citations Generated by AI, THE HILL (Dec. 29, 2023, 1:40 PM), https://thehill.com/regulation/court-battles/4381736-michael-cohengave-lawyer-fraudulent-case-citations-generated-by-ai; Larry Neumeister, Lawyers Submitted Bogus Case Law Created by ChatGPT. A Judge Fined Them \$5,000, ASSOC. PRESS (JULY 22, 2023), https://apnews.com/article/artificial-intelligence-chatgpt-fake-caselawyers-d6ae9fa79d0542db9e1455397aef381c.

⁹⁹ 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/artificialintelligence/federal-judges-start-cracking-down-on-the-use-of-artificialintelligence-in-court-filings.

¹⁰⁰ See Shweta Watwe, Judges Reflect on GenAI Use One Year After ChatGPT's Debut, BLOOMBERG LAW (Nov. 28, 2023, 4:00 AM), https://news.bloomberglaw.com/litigation/judges-reflect-on-genai-useone-year-after-chatgpts-debut.

¹⁰¹ See Megan Cerullo, *Texas Judge Bans Filings Solely Created* by AI After ChatGPT Made Up Cases, MONEYWATCH (June 2, 2023, 2:07 PM), https://www.cbsnews.com/news/texas-judge-bans-chatgpt-courtfiling.

¹⁰² 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) (on file with authors) (speculating that generative AI may significantly impact the quantity and quality of judicial opinions).

50

barriers to justice.¹⁰³ 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 ediscovery 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.¹⁰⁴ 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.¹⁰⁵

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.¹⁰⁶ Historically, shifts in legal

¹⁰³ 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.").

¹⁰⁴ See Jon J. Lee, *Reimagining Attorney Regulation*, B.C. L. REV. (forthcoming 2024) (noting that rules governing attorney misconduct have historically been general in nature).

¹⁰⁵ 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, 3:28 AM), https://www.reuters.com/legal/new-york-lawyers-sanctioned-usingfake-chatgpt-cases-legal-brief-2023-06-22.

¹⁰⁶ See e.g., Joseph Landau & Ron Lazebnik, Law Schools Must Embrace AI, NAT'L L.J. (July 10, 2023, 11:00 AM), https://www.law.com/nationallawjournal/2023/07/10/needs-edit-lawschools-must-embrace-ai; Kristen Baginski & Celeste Pometto DiNicola, AI Goes to Law School, LEXISNEXIS LEGAL INSIGHTS (Dec. 12, 2023), https://www.lexisnexis.com/community/insights/legal/b/thought-

leadership/posts/ai-goes-to-law-school ("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 practitioners. This

technology have only had a limited effect on legal training. This is particularly true when it comes to first-year law students, who have long studied the same mandatory curriculum, which is typically taught to them through some form of Socratic instruction.¹⁰⁷ Although recent decades have seen important adaptations to this approach—from more inclusive Socratic questioning,¹⁰⁸ to an increased focus on statutory interpretation,¹⁰⁹ to increased opportunities for formative feedback¹¹⁰ 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.¹¹¹ 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

¹⁰⁸ 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 "studentcentered, skills-centered, client-centered, and community-centered").

¹⁰⁹ 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).

¹¹⁰ See 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 firstyear law students on mid-term exams improved students' performance in their other first-year classes).

¹¹¹ See, e.g., Cass R. Sunstein, On Analogical Reasoning, 106 HARV. L. REV. 741 (1993) (exploring the distinctive nature of legal reasoning).

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

¹⁰⁷ 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.").

craft legal arguments requires many of the same basic legal and analytical skills as other forms of lawyering, including a capacity to question initial answers, confirm the accuracy of arguments and sources, organize issues clearly, and assess the strength of alternative arguments.¹¹²

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. 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.¹¹³

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.¹¹⁴ 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. Relying instead on honor codes is simply impractical given the current power of widely accessible generative AI tools.¹¹⁵ This is

¹¹² See Schwarcz & Choi, AI Tools for Lawyers, supra note 64.

¹¹³ See Erin McCauliff, The Calculator in the Elementary Classroom: Making a Useful Tool out of an Ineffective Crutch, 21 CONCEPT 1 (2003).

¹¹⁴ See Choi & Schwarcz, AI Assistance in Legal Analysis, supra note 8.

¹¹⁵ See Julianne Hill, Profs Trade Notes as Law Schools Write Generative AI Policies, ABA J. (Jan. 2, 2024, 1:12 PM), https://www.abajournal.com/web/article/law-profs-trade-notes-as-lawschool-write-generative-ai-policies (describing different law schools'

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.¹¹⁶ 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.¹¹⁷

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.¹¹⁸ It is also supported by the differential impact of AI on quality across the different task types; whereas students interested in some practice areas may rightly believe that it would not be a good use of their law school credits to take a class that focuses significant attention on using generative AI, other students may rightly reach the opposite conclusion depending on their career aspirations and interests.

The quantity and scope of these classes should of course vary by school and context, though law schools with students who are more interested in or likely to provide legal services to individuals or costsensitive clients should be particularly aggressive in developing these

¹¹⁸ See supra Part III (describing survey results of participants).

efforts to revise their academic integrity codes in response to generative artificial intelligence).

¹¹⁶ See, e.g., Jiang Zhengyuan, Zhang Jinghuai, & Neil Zhenqiang Gong, Evading Watermark Based Detection of AI-Generated Content (May 5, 2023) (unpublished manuscript) (on file with authors) (discussing the difficulty of detecting AI-generated text even if sophisticated technological techniques for "watermarking" such text is attempted).

¹¹⁷ See Daniel Houser, Stefan Vetter, & Joachim Winter, Fairness and Cheating, 56 EUR. 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. & HUMAN DECISION PROCESSES 157 (2011) ("We cheat because we think others are cheating.").

course offerings. So too should law schools that focus on producing "practice-ready" attorneys who are less likely to receive extensive on-thejob training early in their career.¹¹⁹ 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.¹²⁰

CONCLUSION

We conducted the first randomized controlled trial to evaluate LLM assistance with legal analysis. 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.

¹¹⁹ 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, 150 (2018) (arguing that "students 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").

¹²⁰ See Schwarcz & Choi, AI Tools for Lawyers, supra note 64 (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.¹²¹ 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.¹²² Training was split into three subareas. The first covered general principles on using AI effectively in legal research and writing.¹²³ 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

¹²¹ This training drew heavily on previous work by two of us. *See* Schwarcz & Choi, *AI Tools for Lawyers, supra* note 64.

¹²² 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.

¹²³ These general principles included the following key pieces of advice: (i) Think about any legal problem first-develop your own basic instincts about key issues, principles, and parameters of work product you will need to produce; (ii) Start prompts by giving AI context that it should use to approach a question (i.e. "You are an experienced litigator"); (iii) Use AI to refine initial assessment of project by asking it to produce an outline, identify key issues, or produce first draft (in case of shorter assignments); (iv) Chunk up elements of outline, issues, application of rules into bite-sized bits, and ask AI to analyze each bit, adjusting level of generality based on problem, quality of answers; (v) Provide AI with all the key details that a person would need to accomplish prior step; (vi) Iterate by providing additional details that you may have left out, asking AI to alter elements that do not look good, or asking AI to elaborate on elements that do look promising; (vii) Provide AI with relevant source materials, including cases, statutes, contract parameters, etc.; (viii) Do not rely on AI to conduct specific legal research or identify specific legal source material unless you confirm veracity of that material.

relevant legal rules or source materials within prompts.¹²⁴ Second, the training covered basic techniques for using AI effectively in litigationoriented settings, covering topics such as using AI to summarize and apply primary sources like caselaw and statutes.¹²⁵ 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.¹²⁶

¹²⁴ See Choi & Schwarcz, AI Assistance in Legal Analysis, supra note 8. For a review of the computer science literature on these prompting strategies, see, e.g., Prompt Engineering, OPENAI, https://platform.openai.com/docs/guides/prompt-engineering (last visited Jan. 9, 2024); Alan D. Thompson, Microsoft Bing Chat (Sydney/GPT-4), LIFE ARCHITECT (Feb. 22, 2023), https://lifearchitect.ai/bing-chat; Tyler Cowen & Alexander T. Tabarrok, How to Learn and Teach Economics with Large Language Models, Including GPT (Mar. 17, 2023) (unpublished manuscript) (on file with authors). See also AI and Machine Learning Experts, Experienced Attorneys, Thousands of Hours of Prompt Engineering—and That's Just to Launch, CASETEXT (May 12, 2023), https://casetext.com/blog/building-an-ai-legal-assistant-lawyers-cantrust; Jason Wei et al., Chain-of-Thought Prompting Elicits Reasoning in Large Language Models, in PROCEEDINGS OF THE 36TH INTERNATIONAL **CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS 4356** (2022). Tom B. Brown et al., Language Models Are Few-Shot Learners, in

ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 33 (2020); Baolin Peng et al., Check Your Facts and Try Again: Improving Large Language Models with External Knowledge and Automated Feedback (Mar. 8, 2023) (unpublished manuscript) (on file with authors).

¹²⁵ 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.

¹²⁶ 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, ambiguities in deal terms; (iv) Help issue-spot

57

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:

(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.

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.

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 complaintdrafting 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

LAWYERING IN THE AGE OF AI

distributions were generated by calculating means on bootstrapped distributions, with 10,000 iterations.



Figure 17: Difference in Grade with Access to AI—Complaint Drafting



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

Figure 19: Difference in Grade with Access to AI—Employee Handbook





Figure 20: 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— Employee Handbook





Figure 24: Difference in Time Taken with Access to AI—Client Memo