

**Generative Artificial Intelligence And Access to Justice:
Possibilities, Concerns, Best Practices, and How to Measure Success**
Cas Laskowski, Christopher L. Griffin, Jr.,** & Samuel A. Thumma****

December 1, 2023

I. Introduction

By one definition, “[g]enerative artificial intelligence (AI) describes algorithms (such as ChatGPT) that can be used to create new content, including audio, code, images, text, simulations, and videos.”¹ Any form of AI, whether generative or “general,” meaning capable of processing information just like a human being, used to be the stuff of science fiction. Although we have not created machines as advanced as the HAL 9000 in *2001: A Space Odyssey* (1968), or the human-like child played by Haley Joel Osment in *AI: Artificial Intelligence* (2001), the speed at which AI continues to evolve is staggering.

As noted in February 2023, “[a]fter years of research,” generative AI “is reaching a sort of tipping point, capturing the imaginations of everyone from students saving time on their essay writing to leaders at the world’s largest tech companies. Excitement is building around the possibilities that AI tools unlock, but what exactly these tools are capable of and how they work is still not widely understood.”² Now, ten months later, a Google search for “ChatGPT” reveals about 1.5 billion results. And that is just one of many platforms in the generative AI space.

The ostensible purpose of these technologies is boosting our collective efficiency. Just as the Industrial Revolution heralded the replacement of human labor with automation, the AI-led revolution could save millions of hours of cognitive processing using powerful algorithms. These tools stand to transform any number of vocations, including the legal profession. Attorneys could spend more time on client relations than contract drafting. Courts could identify better ways to help individuals through the legal system and resolving disputes.

Along with the extraordinary potential of generative AI, we should not lose sight of the extraordinary risks. In this paper, we highlight both dimensions in the context of promoting access to justice. By access to justice, we mean any practice that helps litigants—especially in the nation’s civil courts—navigate their legal matters with minimal or no formal attorney representation. We also include efforts that help potential litigants avoid having to invoke the legal system in the first place and how courts and others can improve the legal system to better serve the public. We start with some possibilities, recognizing that generative AI is so new that it

* Head of Research, Data, & Instruction, University of Arizona James E. Rogers College of Law.

** Professor and Director of Empirical & Policy Research, University of Arizona James E. Rogers College of Law.

*** Judge, Arizona Court of Appeals, Division One, Phoenix, Arizona; Chair, Arizona Commission on Access to Justice and Co-chair of the Arizona Supreme Court COVID-19 Continuity of Court Operations During Public Health Emergency Workgroup (aka the Plan B Workgroup). The views expressed do not represent those of the Arizona Courts or the Arizona Court of Appeals.

¹ *What is generative AI?*, McKinsey & Company (Jan. 19, 2023), <https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai> (last visited Oct. 9, 2023).

² *Artificial Intelligence, What is Generative AI? An AI explains*, World Economic Forum (Feb. 6, 2023), <https://www.weforum.org/agenda/2023/02/generative-ai-explain-algorithms-work/> (last visited Oct. 9, 2023).

is hard to provide more than a tabletop exercise of how it might enhance access to justice. We then address concerns about using generative AI to advance access to justice and assist self-represented litigants. Finally, we discuss how to measure success in using generative AI to bridge the justice gap. At the end of the day, great care is needed in using generative AI to enhance access to justice, to ensure its long-term success, and to address a host of valid concerns.

II. Generative AI's Potential to Enhance Access to Justice

Evaluating generative AI and access to justice is difficult, mostly because the underlying technology is comparatively new and rapidly evolving. A simple example proves the point. Nearly three months apart, we asked the publicly available Bing AI search engine³ the following question: “How can artificial intelligence help advance access to justice”? A comparison of the results is telling.

On August 14, 2023, the response was not all that instructive or optimistic.⁴ First noting AI “can help improve access to justice in many ways,” the response provided generalities like “a more responsive justice system”; “augmenting and even replacing lawyers”; and “provid[ing] a more just legal outcome than a human.” That was quickly followed by a caution that technical advances in the law had not made services cheaper and more accessible, largely because of “the law’s apparent impenetrability.” The response ended on a more hopeful tone, suggesting that AI could help provide legal services at a lower cost to a larger number of people for two reasons: “Firstly, it can support the provision of legal services; and secondly, it can replace the role of legal experts. Legal technology that supports justice includes natural language processing (NLP), machine learning and chatbots.” All these observations are, generally speaking, true; they also seem comically simple to anyone who focuses on access to justice.

On November 4, 2023 – almost three months later – an identical search on the same platform was far more instructive and helpful, providing a response that fits in three broad categories.⁵

First, Bing AI responded: “AI can **increase efficiencies** by automating tasks such as document preparation, legal research, and case management. This can reduce the workload and costs for lawyers and courts, and speed up the resolution of legal disputes.” To the extent that machine learning can perform or aid in performing these tasks at a small fraction of the time a human would take, AI has extraordinary potential for saving lawyers and others time and energy. Smart people have declared that some of these tasks will be “resolved/solved in the near term” (if not currently) by generative AI platforms.⁶ But is Bing AI aiming too low? Why stop here?

³ *What is Bing AI Search?*, <https://searchwith.ai/blog/understanding-how-bing-ai-works/#:~:text=Bing%20AI%20search%20is%20a%20search%20engine%20that,technology%20called%20Prometheus%20to%20enhance%20the%20search%20experience> (last visited Oct. 9, 2023).

⁴ The verbatim response appears in the Appendix.

⁵ The verbatim response appears in the Appendix.

⁶ Katherine B. Forrest & Catherine Nyarady, Paul, Weiss, Rifkind, Wharten & Garrison, LLP, *AI and Access to Justice*, PowerPoint presentation at slide 31 (April 19, 2023), <https://www.nycourts.gov/LegacyPDFS/accesstojusticecommission/tc/2023/3A-AI-and-Algorithmic-Bias.pdf> (last visited Oct. 9, 2023).

A more ambitious agenda might involve generative AI helping resolve disputes before they hit a judge's chambers, or even before parties arrive at the courthouse. For example, some courts are turning to court-adjacent online dispute resolution for high-volume civil disputes (e.g., consumer debt).⁷ At the moment, human facilitators must join the chat to help litigants reach pretrial settlement. But what if generative AI could help better facilitate that process in real time? Another holy grail might be providing objective insight on optimal bargaining positions in voluntary alternative dispute resolution systems.

What if generative AI could provide information on how best to *avoid* litigation altogether? Might a sophisticated algorithm sift through tough issues before cases are filed and positions harden? How about helping parties arrive upon a genuinely understandable and objectively fair residential lease agreement before they find themselves in the eviction context? How about using AI to help educate a self-represented party to know how to take action without needing to invoke the legal system? What if AI could help a person know their rights before the legal crisis arrives? Similarly, what if AI could help non-English speakers better understand, in their native tongues, the legal system?

It is a good and helpful thing for lawyers and judges to have more accurate information, to be more efficient, to have better tools to help assess risk and to give more accurate advice leading to more accurate decisions. A lawyer can offer that advice and information in legal settings. But for many parties to life-altering litigation, the possibility of having a lawyer is unlikely. Common examples include unemployment benefits, evictions, consumer debt collection, and family court actions.⁸ One can easily imagine the impact of generative AI emerging most clearly, not in saving lawyers more time on their case, but providing actionable information to self-represented litigants that they otherwise would never receive. When representing clients, a lawyer can provide advice and counsel on strategy and tactics and advise on what to press (and not to press); perhaps generative AI could perform the same functions for people who are not represented by lawyers and have little or no chance of retaining counsel.

Relatedly, and quite powerfully, generative AI might explain to courts and academics why self-represented parties do not use technology (like electronic document filing) that attorneys are required to use.⁹ Can generative AI help determine why self-represented parties are getting childcare, taking time off work, finding transportation, parking and the right courthouse, and filing documents in paper form with a filing clerk when, instead, they could do so using the internet from the comfort of their own home? Can generative AI help identify the best days of

⁷ For one prominent example, see Utah Courts, *Online Dispute Resolution (ODR) Pilot Project*, <https://legacy.utcourts.gov/odr/> (last visited Nov. 29, 2023).

⁸ See Samuel A. Thumma & Jaqueline E. Marzocca, *The Self-Represented Party The Most Unique Party of Them All*, 59 ARIZONA ATTORNEY 24, 26 (June 2023) ("Nationwide, estimates provide that 'more than 70 percent of civil and family cases involve at least one self-represented party.' . . . In Arizona, the percentages may be even higher. For Maricopa County Superior Court cases closed during the 12 months ending June 30, 2021 (FY 2021), more than 90 percent of family court cases had at least one self-represented party, and more than 70 percent of the cases involved both parties being self-represented.") (citations omitted).

⁹ Data on file with the authors show that, from July 1, 2022 to May 31, 2023, of the 446,154 filings by self-represented litigants in family court cases in Maricopa County Superior Court, 432,797 (or 97 percent) were paper filings, while only 13,175 (or 3 percent) were documents that were efiled (which is required by lawyers representing litigants).

the week, and best times of the day, to help ensure parties appear for court hearings?¹⁰ The answers to these questions will, at least in part, show whether generative AI can improve the legal system (in court-based litigation, court-adjacent efforts, or completely outside the court system) for those who otherwise get lost in the legal shuffle.

Second, Bing AI responded: “AI can democratize access to legal information by providing online platforms and tools that can answer legal questions, generate legal documents, and offer guidance and advice. This can help people who cannot afford or access lawyers to solve their own legal problems or connect them with licensed professionals who can.” These are powerful examples of how generative AI can promote access to justice by delivering the law to people’s digital devices on demand. And some wise commentators have advocated such things separately.¹¹ But, again, is Bing AI “thinking” too narrowly?

For all the good that would follow more people getting their hands on legal rules, those rules are often unnecessarily complex. What if generative AI could identify the most problematic bottlenecks in legal processes and simplify them? Could AI see these macro-level patterns that elude even the most intelligent lawyers and make them more user-friendly for self-represented litigants—and even for seasoned lawyers? Procedural rule sets never (or almost never) get shorter and simpler over time. Maybe AI could change that trend and streamline rules so that *everyone* can understand them. Instead of a legal system dominated by centuries-old, arcane, and at times foreign language—or at least a seemingly different dialect of American English—technology might generate rule sets that facilitate rather than frustrate comprehension and use.

Any of these functions inherently makes the law more inclusive and applicable. AI likely won’t replace the three years of formal training that licensed attorneys need to practice, but it can help them express legal arguments more plainly and, hopefully, more effectively.

Third, Bing AI Responded: “AI can improve the quality and consistency of legal decisions by using data and algorithms to analyze cases, predict outcomes, and recommend actions. This can help judges and lawyers to make more informed and objective decisions, and reduce the risk of human errors and biases.” The platform added: “AI can **enhance the transparency and accountability of the justice system** by making legal data and processes more accessible and understandable to the public. This can increase the trust and confidence of the people in the rule of law, and encourage participation and feedback.”

If true, this prediction is powerful. As a system administered by human beings, the justice system is not free from bias or discrimination. And generative AI provides the possibility of

¹⁰ Data on file with the authors show that, from July 12, 2023 through October 31, 2023, of the 4,594 total initial eviction hearings in the Pima County Consolidated Justice Court, the average appearance rate was 61%, but that the appearance rate on Mondays was 60%, while the appearance rate on Tuesdays was 62%.

¹¹ Clare Fraser, *AI: Opening the Door to Justice: How We Can Enhance Access to Justice – and Prevent Inequality – by Developing a Customised Artificial Intelligence Model with the Citizen as the End User*, Law Society of Scotland (August 2023) (advocating for creating “a customised large language model (“LLM”) within an environment where data such as case law, codes of practice and guidance [for the laws of Scotland] have been uploaded and embedded. The LLM is developed with the citizen as the predominant user and not the legal professional.”), <https://www.lawscot.org.uk/members/journal/issues/vol-68-issue-08/ai-opening-the-door-to-justice/> (last visited Oct. 9, 2023).

offering great advances in reducing those flaws. But we ask again: is Bing AI not ambitious enough?

Just as it might lend a hand in rationalizing an overly complex legal system, generative AI might help the justice system determine which metrics are valid and consistent and which are not. Generative AI could be equipped with vision that cuts through longstanding problems, using enormous data sets to identify and help set better standards. For example, many courts and social scientists point to the recidivism rate as an indicator of success in a specialized treatment court. What if they're wrong? What if generative AI could help determine what is the best indicator of success? Similarly, is affording individual judicial officers more, or less, discretion when deciding issues involving liberty (e.g., pretrial detention) the better path? Or are there several avenues, not just one, that generative AI can help identify?

On this dimension, the future is uncertain. Generative AI possesses previously unimaginable capacity. But can generative AI have an influence and be accepted as valid and helpful by the various participants involved in, for example, the criminal justice system? Law enforcement officers, prosecutors, defenders, courts, prisons, and boards of parole and clemency all come to the system with their professional experience and conventional wisdom. For generative AI to break through the inertia, it has to prove its own efficacy by teaching its human users how to look at the world correctly, or at least better.

Generative AI provides a tool, but not a panacea, for addressing time-worn, intractable issues with new and perhaps counterintuitive suggested resolutions. It's time to look hard and deeply at those potential solutions that generative AI makes possible. But in doing so, it is essential to address best practices and recognize concerns generative AI presents, and also have a careful eye focused on how to measure success.

III. Concerns over Using AI and Best Practices

To leverage AI and enhance access to justice, we must understand its limitations and cultivate best practices toward empowering users instead of augmenting inequities.

All machine learning model outputs are no more than information predictions and collections. We refer to the models as "learning" things because they undergo a process designed to mirror the way humans absorb information. AI algorithms are initially "educated" on a set of training data, mapping patterns in those data until it can receive new information and generate accurate connections or identify valid patterns. For example, if we are training an algorithm to perform facial recognition tasks, we might feed it a series of images of people's faces (as well as pictures of other items). The more faces it "sees," the better it can identify what factors are most important to correctly picking out faces "in the crowd."

When they work, these systems are formidable. Understanding their limitations in any particular context (e.g., in criminal investigations) is critical to mitigating the risks of incorrect prediction and ensuring due process in their implementation. In the access to justice realm, inaccurate predictions could be devastating. If self-represented litigants rely on generative AI to navigate civil legal issues, we hope the algorithm will not needlessly lead them astray. In

generative text models, like OpenAI's GPT-4, the answer to a question or prompt is also a prediction: the most likely next word or phrase based on a large language model. As with facial recognition technology, the accuracy and usability of an AI response to a question about handling an eviction case will depend on the quality, scale, and variability of the data on which the algorithm was trained, as well as the structure of the prompt itself.

How instructive any response is to the prompt a person enters will, at first, depend on the data used to train the model. Although it might seem obvious, an inanimate algorithm cannot (at least not yet) learn from information to which it was not exposed. This truth leads to a shortcoming of generative AI known as *exposure bias*. Exposure bias emerges when a computer model trained on a specific set of data 1) does not perform well when introduced to different data and 2) fails to creatively and accurately interpret the new data. This problem literally feeds on itself. In the context of generative AI, exposure bias grows rapidly and perniciously. Any prediction of the next word in a generative response relies on *all the previous words* the model generated. As a result, any error early on compounds the “garbage in-garbage out” dilemma.

A recent cautionary example comes from the “Tessa” generative chatbot used by the National Eating Disorder Association (NEDA). The algorithmic system was designed to replace humans at a call center for people dealing with disordered eating. Because generative AI models must be trained on a wide cross-section of input data to provide sufficient responses, the training data needed to include enough examples of helpful reactions to someone in distress. Unfortunately, *dieting advice* is much more prevalent in everyday human conversation; the training data, therefore, “taught” the algorithm with language more consistent with restricted eating. As a result, by early June 2023, NEDA had suspended its AI chatbot for giving harmful advice. NEDA’s chief executive, Elizabeth Thompson told the *New York Times* she was “waiting for an explanation about how that content was introduced into a closed program.”¹²

The lesson for access to justice supporters is that potential AI tools must be trained on data reflecting the legal problems facing people across socioeconomic, educational, and geographic distributions—not just the average or higher. This is particularly true for racial and ethnic minorities; their experiences are not recognized by the underlying algorithm, leading to serious errors in advice or decision-making.¹³ At the very least, AI tools designed for self-represented litigants should oversample the cases and circumstances that those individuals most frequently encounter. Otherwise, they might be worse off than without the technology, as in the NEDA example above.

Generative AI is also subject to *hallucinations*, which are inaccurate sentences or phrases produced by the system. While there are methods to reduce such risk, no technique exists to completely eliminate it. Generative AI models create content by predicting the next object in some sequence that is probabilistically ideal, be it a word or a pixel. More advanced generative

¹² Lauren McCarthy, *A Wellness Chatbot Is Offline After Its ‘Harmful’ Focus on Weight Loss*, N.Y. TIMES, <https://www.nytimes.com/2023/06/08/us/ai-chatbot-tessa-eating-disorders-association.html> (June 8, 2023; updated June 9, 2023).

¹³ See, e.g., Brianna Rauenzahn, Jamison Chung & Aaron Kaufman, *Facing Bias in Facial Recognition Technology*, THE REG. REV., Mar. 20, 2021, <https://www.theregreview.org/2021/03/20/saturday-seminar-facing-bias-in-facial-recognition-technology/>

models produce much better prediction outputs but might be cost-prohibitive for access to justice.

In addition to fictitious sentences, well-known hallucinations include generating false citations. False citations arise when the algorithmic model is designed to predict the right combination of words and numbers that mirrors the structure of citations from training data, *without regard to its truth*. A now infamous example involved two plaintiff's attorneys who used ChatGPT to write a legal brief. The AI platform hallucinated six case citations in the document, which defense counsel could not locate in actual reporters.¹⁴ The court ended up dismissing the case and sanctioning the plaintiff's attorneys. In the sanctions order, the court said "that there is nothing inherently improper in lawyers using AI 'for assistance,' but he said lawyer ethics rules 'impose a gatekeeping role on attorneys to ensure the accuracy of their filings.'" ¹⁵

Even worse, some models have been trained to produce real citations but might still apply them incorrectly or look to true citations that are not the best choice for the proposition stated. For example, when we asked a prototype legal chatbot "Can a school prevent a student article from being printed in a school publication?", it responded, in part: "[S]chool authorities can exercise prior restraint on publications distributed on school premises during school hours if they can reasonably forecast substantial disruption of or material interference with school activities due to the distribution of such printed material USCS Const. Amend. 1, Religious and political freedom." Although the answer does follow from a required First Amendment analysis, the better source for citation purposes is the actual United States Supreme Court decision.¹⁶ Now imagine a self-represented litigant using a chatbot to create a pleading or other court document. Without the first clue about how to verify a citation's accuracy, the litigant could wind up submitting completely fictitious information and drawing the court's ire.

Another concern for AI-informed access to justice is the transparency of algorithmic processes. Due process is founded on notice, the opportunity to be heard at a meaningful time in a meaningful way, and the chance to challenge evidence offered against a party.¹⁷ Many AI systems are not capable of providing the reasoning behind their outputs. Without a clear understanding of the factors involved in systems, the bases for a decision, or the ability to challenge it after the fact, due process is at risk.¹⁸

¹⁴ See Larry Neumeister, *Lawyers Submitted Bogus Case Law Created by ChatGPT. A Judge Fined Them \$5,000*, AP (June 22, 2023), <https://apnews.com/article/artificial-intelligence-chatgpt-fake-case-lawyers-d6ae9fa79d0542db9e1455397aef381c>.

¹⁵ Sara Merken, *New York Lawyers Sanctioned for Using Fake ChatGPT Cases in Legal Brief*, REUTERS (June 26, 2023), <https://www.reuters.com/legal/new-york-lawyers-sanctioned-using-fake-chatgpt-cases-legal-brief-2023-06-22/>.

¹⁶ See *Hazelwood Sch. Dist. v. Kuhlmeier*, 484 U.S. 260 (1988); see also Mehul Bhattacharyya et al., *High Rates of Fabricated and Inaccurate References in ChatGPT-Generated Medical Content*, 15 CUREUS, May 19, 2023, at 2, 4. (showing that, when ChatGPT-3.5 was asked "to generate 30 unique short papers on various biomedical topics," the platform listed 115 total references, of which "47% were fabricated, 46% were authentic but inaccurate, and only 7% were authentic and accurate").

¹⁷ See, e.g., *Armstrong v. Manzo*, 380 U.S. 545, 552 (1965); *Grannis v. Ordean*, 234 U.S. 385, 394 (1914).

¹⁸ *But see State v. Loomis*, 881 N.W.2d 749, 761 (Wisc. 2016) (finding that "[the defendant] is correct that the risk scores do not explain how the COMPAS program uses information to calculate the risk scores" but declining to find a due process violation because "to the extent that [the defendant's] risk assessment is based upon his answers to

Best practices for legal AI systems are essential and should embrace, among other things, the following guidelines:

- **Use Diverse, Representative Data.** Bias in AI outputs often stems from biased training data.¹⁹ Ensuring that training datasets include diversity across characteristics that matter for access to justice (e.g., race, ethnicity, income, education) and are representative of the population the model will serve is crucial. This involves including a wide range of demographics, perspectives, and scenarios in the data. When sufficient data are not available, one can collaborate with data scientists who apply technical strategies for reducing bias or improve data collection methods for future analysis. Similarly, testing prototypes with the populations who often represent themselves in court can inform the development process and help identify potential areas of bias. Too often, innovation occurs without the input of critical stakeholders. In the access to justice context, that means testing academic and practitioner assumptions against the lived experience and needs of the target audience.²⁰
- **Create “Human-in-the-Loop” Systems.** Human oversight in AI decision-making processes must be included in any algorithmic platform, especially in a domain like high-volume, high-stakes civil litigation. Keeping people “in the loop” will not guarantee success, but timely human intervention can override decisions that the AI system does not “understand” will be detrimental to users. The level of human oversight needed as well as the timing of oversight depends on the level of risk involved and the potential implications of delay.
- **Develop Impact Assessments.** AI models that courts and lawyers deploy should be reviewed regularly to ensure that the outcomes they expect align with the outcomes they observe as much as possible. When they do not, developers should refine the model to account for those unexpected outputs and to incorporate new data and changing societal norms, both of which reduce bias over time. For example, an impact assessment might have flagged the issues with NEDA’s Tessa chatbot before it caused any harm, potentially flagging how poorly the system responded to people with eating disorders.
- **Be as Transparent as Possible.** Stakeholders should strive for a transparent explanation of how the AI model was developed and how it works so that users can see what factors informed any decision-making process and how they were weighted. Being open about algorithmic inputs and calculations builds trust and understanding among everyone involved in the civil justice system. Educating users about the capabilities and limitations of AI models, as well as providing clear guidelines on how to use these models effectively and responsibly can help mitigate risks.

questions and publicly available data about his criminal history, [the defendant] had the opportunity to verify that the questions and answers listed on the COMPAS report were accurate”).

¹⁹ See, e.g., Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218 (2019).

²⁰ A recent example is STACY BUTLER, SARAH MAUET, CHRISTOPHER L. GRIFFIN, JR. & MACKENZIE PISH, THE UTAH ONLINE DISPUTE RESOLUTION PLATFORM: A USABILITY EVALUATION AND REPORT (2020), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3696105.

IV. How to Measure Success

As the justice system continues grappling with the questions and caveats in the previous two sections, we should also ask how best to measure generative AI's benefits. Legal academics and social scientists now have at their disposal a variety of tools for evaluating new ideas.²¹ A complete review of those methods is beyond the scope of this paper. For now, we highlight some key criteria in choosing how to evaluate AI systems in the access to justice context.

The first consideration is *what* to evaluate. This question is one of outcomes. If we want to know whether an AI tool promotes inclusivity and transparency, we might focus on user comprehension of how the tool works. If we care more about whether self-represented litigants successfully resolve their legal matters, we will choose “win rates” as the relevant outcome variable. And if we want to understand better how AI promotes efficient resolution, we might use time to disposition as the main indicator.

In some sense, there is no “right” choice when it comes to outcome variables. What matters to the empirical analysis is what matters to the community deploying the AI tool. As a result, measuring success is somewhat in the eye of the beholder. The outcome variables included in any evaluation should reflect the values and needs of those administering and using algorithms. For example, a jurisdiction that wants its online dispute resolution tool to be useful without relying too much on human technical support would care a lot about whether users can find answers in the frequently asked questions (FAQ) section.²² But constantly turning to the FAQ can also signal that the platform is non-intuitive or too cumbersome to follow.

The second consideration that courts and administrators should confront is *how* to evaluate. Again, there are many more evaluative methods from applied statistics than there is space to review here. Suffice it to say that there are three primary approaches: (1) subjective surveys; (2) observational data; and (3) experimental methods. Surveys, by construction, can only reveal (if anything) how and why users interact (or don't) with a legal innovation. They can be informative about efficacy—insofar as user satisfaction measures how well something works—but they are usually unable to say much objectively about outcomes. Observational studies rely on large datasets capturing measures of the chosen outcome (the dependent variable) and all the discernable factors that could plausibly impact it (the independent variables). The social scientist using observational methods often wants to generate causal inference. They can't, however, because the processes that created the data are subject to selection effects and other “confounding” influences. For example, look at a circumstance where a court deploys an AI dispute resolution platform and, when used, it resolves cases more quickly than when it was not. That result *could* reflect the utility of the AI tool. It could also pick up the unobserved impact of inherent diligence if the people who choose to use the tool get things done more quickly (on average) than those who choose not to use the tool. At the extreme, the tool could be useless, and the data only produce the results because users are faster in general than non-users in taking care of business.

²¹ A very accessible overview appears at Stanford Legal Design Lab, *Evaluation Methods of Justice Innovations*, <https://justiceinnovation.law.stanford.edu/resources/evaluation/> (last visited Nov. 30, 2023).

²² See BUTLER, ET AL., *supra* note 2020, at 35-39.

The gold standard methodology for assessing whether any legal innovation, AI-based or otherwise, is the randomized control trial (RCT). Simply stated, RCTs follow the procedure of a clinical trial: participants are divided into a control group and one or more treated groups using some randomizing device (e.g., coin flip, wheel spin). The treated groups are exposed to the innovation, and the control group is shielded from the innovation as much as possible. Allocating the new tool or resource ensures (on average) that any selection effects or confounds will wash out in the analysis. Experimentation along these lines is anathema to many jurists and lawyers. As such, the legal profession lags behind others in the evidence basis for its practices.²³ But if we really want to learn what works and what doesn't—if we want to uncover causality in legal process—we should embrace experimental methods more readily.

Finally, stakeholders must ask *when* to evaluate AI tools for access to justice. This question might be the most practically important for courts and litigants. If an AI-related innovation enters public spaces before being subjected to rigorous testing, any of the terrible consequences outlined in the previous section could accrue. If so, the stakeholders involved would have to admit that they deployed a new procedure without fully understanding (or understanding at all) the likely effects. Even with relatively benign interventions like self-help materials in courthouses, failure to evaluate beforehand risks all sorts of unintended consequences. Thus, justice system stakeholders should at all costs avoid launching AI platforms at scale without a proper *ex ante* evaluation.

The double gold standard path forward, as it were, would be to pilot an AI-backed tool with a small, but statistically powerful, number of users in an RCT. Doing so both provides preliminary evidence of whether the tool works and helps developers weed out bugs. One court in a state could be the pilot jurisdiction, or one courtroom in a county. Armed with solid evidence of effectiveness, administrators could refine the effort and scale up the pilot to more locations and repeat the evaluation. Repeated findings that the AI platform works offers a proper evidentiary basis for full deployment. Aside from this admittedly time-consuming process, courts and lawyers should at least pursue rigorous evaluation no later than the tool's launch date. Along with the downsides mentioned above, officers of the law are reluctant to abandon practices they *believe* are useful. The more entrenched an innovative practice becomes over time, the harder it may be to discard—even if later evaluation shows that it is (and perhaps never was) useful.

V. Conclusion

Generative AI is opening doors to rooms that, until very recently, we didn't even know existed and could not imagine. Its capacity for processing all the information in the country's law libraries and more has enormous potential for enhancing access to justice. The most commonly used chatbot today provides decent answers to the question that we set out to answer in this paper. But those answers are incomplete. The AI platform is less ambitious, does not appreciate the significant risks for self-represented litigants, and has cannot self-assess its responses. These truths reinforce the great care needed when using generative AI to enhance access to justice, to ensure its long-term success, and to address a host of valid concerns. In a very real way, in the future, the sky is the limit—provided we understand generative AI's promise and pitfalls now.

²³ See, e.g., D. James Greiner & Andrea Matthews, *Randomized Control Trials in the United States Legal Profession*, 12 ANN. REV. L. & SOC. SCI. 295 (2016).

Appendix

Bing AI Search Engine (<https://www.bing.com>) responses to the following query on different dates: “How can artificial intelligence help advance access to justice.”

August 14, 2023 at 3:47 pm (Arizona time) search = 220 words and two sources:

Artificial intelligence (AI) can help improve access to justice in many ways. According to the Centre for Public Impact (CPI), AI can provide a more responsive justice system by augmenting and even replacing lawyers. Intelligent machines can help provide a more just legal outcome than a human¹. In many areas of human activity, technological advances have made goods and services cheaper and more accessible. Thus far, however, in law (as in medicine) technological advances have tended to have the opposite effect. A major contributor to this fact is the law’s apparent impenetrability. The law is perceived as a complicated organism that requires the costly intervention of legal experts to be accurately deciphered. Accordingly, even as our technology gets smarter and faster, the trend in the law has been for legal services to get bigger and more costly. But will this always be the case? Intelligent machines now increasingly shape our society. We routinely rely on algorithms to perform tasks, often with greater precision and reliability than humans¹.

AI can provide a legal service at a significantly reduced unit cost to a much larger portion of the population. Firstly, it can support the provision of legal services; and secondly, it can replace the role of legal experts. Legal technology that supports justice includes natural language processing (NLP), machine learning and chatbots².

1. <https://www.centreforpublicimpact.org/insights/joel-tito-ai-justice>
2. <https://www.abajournal.com/magazine/article/artificial-intelligence-has-made-great-inroads-but-not-as-far-as-increasing-access-to-civil-justice>

November 4, 2023 at 7:44 am (Arizona time) search = 587 words and eight sources:

Artificial intelligence (AI) is a powerful technology that can help advance access to justice in many ways. Some of the possible benefits of AI for the justice system are:

- AI can **increase efficiencies** by automating tasks such as document preparation, legal research, and case management. This can reduce the workload and costs for lawyers and courts, and speed up the resolution of legal disputes.
- AI can **democratize access to legal information** by providing online platforms and tools that can answer legal questions, generate legal documents, and offer guidance and advice. This can help people who cannot afford or access lawyers to solve their own legal problems or connect them with licensed professionals who can.

- AI can **improve the quality and consistency of legal decisions** by using data and algorithms to analyze cases, predict outcomes, and recommend actions. This can help judges and lawyers to make more informed and objective decisions, and reduce the risk of human errors and biases.
- AI can **enhance the transparency and accountability of the justice system** by making legal data and processes more accessible and understandable to the public. This can increase the trust and confidence of the people in the rule of law, and encourage participation and feedback.

However, AI also poses some challenges and risks for the justice system, such as:

- AI can **create ethical and legal dilemmas** by raising questions about the responsibility, liability, and accountability of the developers and users of AI systems. For example, who should be held accountable if an AI system makes a wrong or harmful decision, and how can the affected parties seek redress?
- AI can **exacerbate existing inequalities and injustices** by reflecting and reinforcing the biases and discrimination that exist in the data and algorithms used by AI systems. For example, an AI system that predicts the risk of recidivism for criminal defendants may be influenced by factors such as race, gender, and socioeconomic status, and result in unfair and discriminatory outcomes.
- AI can **undermine the human dignity and autonomy of the people** by reducing their involvement and control over their own legal affairs. For example, an AI system that provides legal advice may not be able to capture the nuances and emotions of the human situation, and may not respect the preferences and values of the people.

Therefore, it is important to ensure that AI is used in a way that respects the principles and values of the justice system, such as fairness, equality, impartiality, and human rights. Some of the possible measures to achieve this are:

- Developing and implementing **ethical and legal standards and guidelines** for the design, development, and use of AI systems in the justice system. These should ensure that AI systems are transparent, explainable, accountable, and respectful of human dignity and autonomy.
- Establishing and enforcing **oversight and regulation mechanisms** for the monitoring, evaluation, and auditing of AI systems in the justice system. These should ensure that AI systems are reliable, accurate, consistent, and compliant with the ethical and legal standards and guidelines.
- Providing and promoting **education and training** for the developers, users, and beneficiaries of AI systems in the justice system. These should ensure that they have the necessary skills, knowledge, and awareness to use AI systems effectively, responsibly, and ethically.

Laskowski, Griffin & Thumma (December 1, 2023)

Please do not reproduce or share without the authors' permission.

AI has the potential to transform the justice system and improve access to justice for many people. However, it also requires careful and thoughtful consideration of its implications and impacts, and the adoption of appropriate measures to ensure that it is used for good and not for evil.

Learn more:

1. papers.ssrn.com
2. bing.com
3. jolt.law.harvard.edu
4. centreforpublicimpact.org
5. lawtimesnews.com
6. yjolt.org
7. capgemini.com
8. papers.ssrn.com