

How to harness AI for justice^{*}



A PRELIMINARY AGENDA FOR USING GENERATIVE AI TO IMPROVE ACCESS TO JUSTICE

BY CHRISTOPHER L. GRIFFIN, JR., CAS LASKOWSKI & SAMUEL A. THUMMA

The artificial intelligence (AI) explosion has reached the legal profession. In particular, *generative* AI, which “describes algorithms (such as ChatGPT) that can be used to create new content, including audio, code, images, text, simulations, and videos,”¹ has become increasingly relevant. 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 are still not widely understood.”² Now, a year later, a Google search for “ChatGPT” generates 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 to enhance our collective efficiency. Just as the Industrial Revolution heralded the replacement

of human labor with automation, an AI-led transformation using powerful algorithms could save millions of hours of cognitive processing time. These tools are poised to transform any number of vocations, including the legal profession.³ Attorneys could spend more time on client relations than contract drafting. Courts could identify better ways to help individuals through the legal system and resolve disputes. Self-represented litigants could navigate some legal problems without having to pay for an attorney. However, along with the extraordinary potential of generative AI, we should not lose sight of the extraordinary risks it poses.

Here, we — a law professor, a law librarian, and a judge — 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, resolve 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 as well as ways in which courts and other stakeholders can improve the legal system to better serve the public.

We start by outlining generative AI’s most promising features, recognizing that generative AI is so new that

it is hard to offer more than a table-top exercise of how it might enhance access to justice. We then address concerns about using generative AI to advance such access and assist self-represented litigants. Finally, we discuss how to measure the success of 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, ensure its long-term success, and address a host of valid concerns.

GENERATIVE AI’S POTENTIAL TO ENHANCE ACCESS TO JUSTICE

Predicting how generative AI will affect access to justice is difficult, mostly because the underlying technology is comparatively new and rapidly evolving. A simple example proves the point: We asked the publicly available Bing AI search engine⁴ the following question: “How can artificial intelligence help advance access to justice”? Then, three months later, we asked it the same thing. A comparison of the results shows just how quickly AI is amassing data.

On August 14, 2023, the response was not all that instructive or optimistic.⁵ First noting that AI “can help improve access to justice in many ways,” the Bing AI provided generalities like “a more responsive justice system”; “augmenting and even ►



To our knowledge, all technologies currently used in ODR platforms require some human facilitator to help litigants reach pretrial settlement. Those humans are usually available in chat spaces or by individual email messages to help the parties reach a mutually acceptable plan. What if generative AI could better facilitate that process in real time?

replacing lawyers”; and “provid[ing] a more just legal outcome than a human.” These responses are fairly standard reflections of what access to justice is supposed to deliver, with or without generative AI. The initial response was quickly followed by a proviso 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, but they also seem comically simple to anyone who devotes their scholarly or practice-related time to access to justice.

On November 4, 2023 — almost three months later — we replicated the search with the identical prompt on the same platform. What followed was far more instructive and helpful, providing a response that maps broadly onto three categories.⁶ We describe these categories below, then suggest how AI might go, helpfully, even further.

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.”⁷

To the extent that machine learning can aid in performing (or outright perform) these tasks at a small fraction of the time a human would expend, AI has extraordinary potential for directing scarce resources toward more complex needs. Some extremely capable minds have declared that such tasks will be “resolved/solved in the near term” (if not currently) by generative AI platforms.⁸ But is Bing AI aiming too low in identifying routine lawyer tasks? Why stop there?

Avoiding Litigation. A truly ambitious agenda might take an even more prophylactic approach, aiding with litigation before the lawyer begins their work, before a complaint reaches the court clerk’s window, or even before parties arrive at the courthouse. For example, some courts are turning to court-adjacent online dispute resolution (ODR) for high-volume civil disputes (e.g., consumer debt).⁹ To our knowledge, all technologies currently used in ODR platforms require some human facilitator to help litigants reach pretrial settlement. Those humans are

usually available in chat spaces or by individual email messages to help the parties reach a mutually acceptable plan.¹⁰ What if generative AI could better facilitate that process in real time? Not only would courts save on human labor costs, but a well-designed algorithm should also be able to narrow the settlement space more accurately and quickly than even the most seasoned mediator.

Avoiding Conflict. As we up the aspirational ante, we might also expect generative AI to guide parties on how best to avoid litigation — or even conflict — altogether. Might a sophisticated algorithm sift through and diagnose difficult issues before suits are filed and positions begin to calcify? For example, how about helping parties arrive at a genuinely understandable and objectively fair residential lease agreement that precludes the need for many summary eviction cases? Could AI help educate a self-represented party on how to solve a problem they face without having to identify it as a legal issue? In addition to increasing the rate of dispute resolution, AI could assist with procedural engagement along the way. For example, generative AI might have the capacity to accurately translate materials and proceedings for non-English speakers in ways that promote procedural fairness

alongside happiness with substantive outcomes.

Providing Legal Advice. It is a good and helpful thing for lawyers and judges to have more accurate information, to be more efficient, to have better tools for assessing risk, and to deliver more actionable advice. But for many parties to life-altering litigation, the possibility of having a lawyer to perform all those functions is unlikely, if not impossible. Common examples include unemployment benefits claimants, tenants in eviction suits, people experiencing consumer or medical debt, and family members arguing over custody or support arrangements.¹¹ One can easily imagine the impact of generative AI most clearly not in saving lawyers more time on their case, but by providing comprehensible information to self-represented litigants that they otherwise would never receive. When representing clients, a lawyer usually provides strategic advice and counsel and suggests which issues should be litigated fiercely (and which should not). 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.

Streamlining the Court Experience. Relatedly, and quite powerfully, generative AI might help courts and academics understand why self-represented parties eschew technology (e.g., electronic document filing) that attorneys are required to use.¹² Can generative AI shed light on why self-represented parties are obtaining childcare, taking time off work, finding transportation, walking through courthouse halls, and filing hard-copy documents when, instead, they could handle pleadings over the internet from the comfort of their own homes? Can generative AI help identify the best

days of the week, and best times of the day, to help ensure parties appear for court hearings?¹³ The answers to these questions will, at least in part, reveal 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 formidable examples of how generative AI can promote access to justice by delivering the law on demand to people’s digital devices. Some wise commentators have advocated such advancements.¹⁴ But, again, is Bing AI “thinking” too narrowly?

Simplifying the Law. For all the good that might follow from more people directly using legal rules, the law’s text and structure are often unnecessarily complex. What if generative AI could identify the most problematic bottlenecks in legal processes and simplify them? For example, if self-represented plaintiffs routinely find their claims dismissed, we might look first to service of process rules. Generative AI could sort among the many reasons why these plaintiffs fail to serve: Is the culprit the limited methods available, the time limits, something else? AI might be able to detect macro-level patterns that elude even the most intelligent lawyers and spur reforms that make rules more user-friendly for self-represented litigants — and even for seasoned lawyers. Procedural rules

never (or almost never) get shorter and simpler over time. Maybe AI could presage a reversal of 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 rules that facilitate rather than frustrate their use. Any of these functions inherently makes the law more inclusive and applicable. And, similarly, AI may help attorneys write and present their own 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 would be monumental. As a system administered by human beings, the justice system is not free from bias or discrimination. And generative AI provides the possibility of offering great advances in reducing those flaws. But we ask again: Is Bing AI not ambitious enough?

Improving Decision-making. Just as it might lend a hand in rationalizing an overly complex legal system, generative AI might help the system determine which justice indicators are valid and which are not. Generative AI could supply a macro-level vision ►

to cure longstanding problems, using enormous datasets to identify and help set better standards. For example, many courts and social scientists suggest that a specialized drug-treatment court should produce lower recidivism rates to be considered successful. But even how recidivism is measured (if recidivism is the right metric) depends not only on the goals of the community but also on the definition of the relevant offense (arrest for another drug crime, or something else?) and the relevant time period (during treatment, one month thereafter, five years thereafter?). Generative AI might help courts parse through complex datasets and select the best indicator of success, conditional on a jurisdiction's resources, values, and objectives. Generative AI could also help answer questions about the ideal amount of judicial oversight as well as the optimal amount of discretion in, for instance, pretrial release or sentencing conditions.

On this dimension, the future is decidedly uncertain. Generative AI's influence will extend only as far as stakeholders accept its results as valid. That may well depend upon what kinds of datasets courts retain and make available to researchers, issues that implicate choices like which electronic case management systems to use, and policy issues including what court data are disclosed. It may also depend on things like a willingness to accept evidence- and data-based changes and improvements. 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 human users how to look at the world in a different way.

Generative AI provides a tool, but not a panacea, for addressing time-worn, intractable issues with new and perhaps counterintuitive solutions. 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, with a careful eye on how to measure success.

POTENTIAL AI PITFALLS

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

Machine learning model outputs are no more than information collections and predictions. We describe these 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 they can receive new information and generate accurate connections or identify valid patterns.¹⁵ 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."¹⁶

When they work, these systems are truly impressive. Understanding their limitations in any particular context (e.g., aiding 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 gen-

erative AI to navigate civil legal issues, incorrect guidance on answering a lawsuit could lead to a default judgment. 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.

Data Inputs. The utility of any response 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 has not been exposed. This truth leads to a shortcoming of generative AI known as exposure bias.¹⁷ 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.¹⁸ This is a problem for generative models because generated text becomes part of the underlying data used to make the next prediction. So a poorly or erroneously generated first sentence will negatively affect the next prediction exponentially.

A recent cautionary example comes from the "Tessa" generative chatbot used by the National Eating Disorder Association (NEDA). Tessa 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 data to provide sufficient responses, those training data needed to include enough examples of helpful reactions to someone in distress. Unfortunately, the



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 human users how to look at the world in a different way.

universe of inputs for NEDA somehow came to include typical human conversation about dieting advice that would not necessarily be appropriate for a population experiencing disordered eating. The training data, therefore, “taught” the algorithm to use language more consistent with restricted eating. As a result, by early June 2023, NEDA had suspended Tessa 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.”¹⁹

The lesson for access to justice advocates is that AI tools must be trained on data reflecting the legal problems facing people across socioeconomic, educational, and geographic distributions — not just the average or endpoints of the distribution. This is particularly true for racial and ethnic minorities; their experiences might not be recognized by the algorithm because of training-data limitations, leading to serious errors in advice or decision-making.²⁰ 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.

Hallucinations. Generative AI is also subject to hallucinations, which are inaccurate sentences or phrases produced by the system.²¹ While there are methods to reduce such risk, no technique exists to completely eliminate it.²² To be sure, more advanced generative models produce much better prediction outputs, but they may be cost-prohibitive for adoption in access-to-justice spaces.

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 the truth*. A now-infamous example involved two plaintiff 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.²³ The court ended up dismissing the case and sanctioning the 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.’”²⁴

Even worse, some models have been trained to produce real citations but

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 may follow from a First Amendment analysis, the better source for citation purposes is the actual United States Supreme Court decision.²⁵ Now imagine a self-represented litigant using a chatbot to draft a pleading or other court document. Without the first clue about how to verify a citation’s accuracy, the litigant could wind up submitting subpar — or perhaps completely fabricated information — and drawing the court’s ire.

Transparency. 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 ►



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 the intended user community. In the access-to-justice context, that means testing academic and practitioner assumptions against the lived experience and needs of the target audience.

against a party.²⁶ Many AI systems are not capable of providing the reasoning behind their outputs.²⁷ And, in some cases, AI creators may hesitate to share their proprietary algorithmic information anyway. 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 imperiled.

BEST PRACTICES

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

- **Use Diverse, Representative Data.** Bias in AI outputs often stems from biased training data.²⁸ Ensuring that training datasets reflect diversity across the many dimensions that matter for access to justice (e.g., race, ethnicity, income, education) is crucial. Without a wide range of demographics, perspectives, and scenarios in the data, any AI tool will surely underserve its intended user base. When representative data are not available, data scientists can apply technical strategies for reducing bias or improving data-collection methods for future analysis. Similarly, testing proto-

types 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 the intended user community. In the access-to-justice context, that means testing academic and practitioner assumptions against the lived experience and needs of the target audience.²⁹

- **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, to the extent possi-

ble. When they do not, developers should refine the model to address those unexpected outputs and to incorporate new data and changing societal norms, both of which can reduce bias over time. For example, impact assessments can flag issues (like those that arose with NEDA’s Tessa chatbot) before they cause any harm by identifying how the system responds.³⁰

- **Be as Transparent as Possible.** Stakeholders should strive for transparent explanations of how any AI model was developed, and how it works, so that users can see what factors informed the 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 them effectively and responsibly, can help mitigate risks.

HOW TO MEASURE SUCCESS

As the justice system grapples with these questions and caveats, it should simultaneously deploy a suite of evaluation tools for measuring generative

AI's benefits. Legal academics and social scientists now have at their disposal a variety of methodologies for program evaluation.³¹ A complete review of those methods is beyond the scope of this essay. For now, we highlight some key criteria for evaluating AI systems in the access-to-justice context.

What to Evaluate. 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 want to know 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 dispute 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. Thus, 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 algorithmic systems. 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.³² But constantly turning to the FAQ can also signal that the platform is nonintuitive or too cumbersome to follow.

How to Evaluate. The second consideration that courts and administrators should confront is *how* to evaluate. Again, there are many more evaluative methods from applied statistics

than 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 must be combined with more objective data to tell a complete story.³³

Observational studies rely on large datasets, including 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 find evidence consistent with causal inference. They often can't, however, because the processes that created the data are subject to selection effects and other “confounding” influences.³⁴ For example, consider a court that deploys an AI dispute-resolution platform that, when used, resolves cases more quickly than the status quo ante. 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 result only due to the fact that users are faster workers in general than nonusers.

The gold standard methodology for assessing any legal innovation, AI-based or otherwise, is the randomized control trial (RCT). In short, RCTs follow the procedure of a clinical trial: Participants are divided into a control group and one or more treated groups, determined by 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.³⁵ Experimentation along these lines — which might exclude some participants due to random assignment or other methodological approaches — is anathema to many jurists and lawyers because they prefer to allocate their time and talent, for example, based on perceived merit. As such, the legal profession lags behind others in the evidence basis for its practices.³⁶ But if we really want to learn what works and what doesn't — if we want to begin to uncover causality in legal process — we should embrace experimental methods more readily.

When to Evaluate. 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 goes public before being subjected to rigorous testing, any of the adverse consequences previously outlined could accrue. If so, the stakeholders involved would have to admit that they deployed a new procedure without fully understanding (or understanding at all) its 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 findings from a proper evaluation in hand.

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 pre- ►

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 the intended user community. In the access-to-justice context, that means testing academic and practitioner assumptions against the lived experience and needs of the target audience.

liminary evidence of whether the tool works and helps developers weed out bugs. One court in a state, or one courtroom in a county, could be the pilot jurisdiction. 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 offer a proper evidentiary basis for full deployment. If this iterative process would be too costly or take too long, courts and lawyers should at least pursue rigorous evaluation at the same time they introduce the innovation in practice. 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.

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 jus-

tice. The most commonly used chatbot today provides decent answers to the question that we set out to answer in this essay. But those answers are incomplete. The AI platform fails to comprehend its true potential as well as its risks, especially for self-represented litigants. 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. At the risk of hyperbole, in the future, the sky is the limit — provided we understand generative AI's promise and pitfalls now.



CHRISTOPHER L. GRIFFIN, JR. is a professor and director of empirical and policy research at the University of Arizona James E. Rogers College of Law.



CAS LASKOWSKI is head of research, data, and instruction at the University of Arizona James E. Rogers College of Law.



SAMUEL A. THUMMA is a judge of the Arizona Court of Appeals, Division One, in Phoenix, Arizona. He serves as chair of the 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 are his own and 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 ECON. FORUM (Feb. 6, 2023), <https://www.weforum.org/agenda/2023/02/generative-ai-explain-algorithms-work> (last visited Oct. 9, 2023).
- ³ See, e.g., John Villaseñor, *How AI Will Revolutionize the Practice of Law*, BROOKINGS (Mar. 20, 2023), <https://www.brookings.edu/articles/how-ai-will-revolutionize-the-practice-of-law>.
- ⁴ Timo Bakker, *Understanding How Bing AI Works*, SEARCH WITH AI (last visited Oct. 9, 2023), <https://searchwith.ai/blog/understanding-how-bing-ai-works>.
- ⁵ The verbatim response appears in the Appendix, which is available with this article at <http://www.judicature.duke.edu>.
- ⁶ The verbatim response appears in the Appendix, which is available with this article at <http://www.judicature.duke.edu>.

- 7 All bolded text in Bing AI responses reflects emphasis in the original.
- 8 KATHERINE B. FORREST & CATHERINE NYARADY, AI AND ACCESS TO JUSTICE 31 (2023), available at <https://www.nycourts.gov/LegacyPDFS/accesstojusticecommission/tc/2023/3A-AI-and-Algorithmic-Bias.pdf>.
- 9 For one prominent example, see *Online Dispute Resolution (ODR) Pilot Project*, UTAH COURTS (last visited Nov. 29, 2023), <https://legacy.utcourts.gov/odr>.
- 10 See, e.g., Melissa Stiglich, *Utah Online Dispute Resolution Pilot Project*, NAT'L. CTR. STATE CT. (Dec. 2017), <https://cdm16501.contentdm.oclc.org/digital/collection/adr/id/63> (describing the human facilitator's responsibilities in Utah's ODR process); Joint Technology Committee, *Resource Bulletin: ODR for Courts*, NAT'L. CTR. STATE CT. (Nov. 29, 2017), https://www.ncsc.org/_data/assets/pdf_file/0031/18499/2017-12-18-odr-for-courts-v2-final.pdf (anticipating the need and potential role of facilitators in state courts); Information Technology Advisory Committee, *Online Dispute Resolution (ODR) Workstream Findings & Recommendations*, JUD. COUNCIL CA. (June 24, 2021), https://www.courts.ca.gov/documents/ODR_Workstream_Report.pdf (outlining the duties of facilitators in California's courts).
- 11 See Samuel A. Thumma & Jaqueline E. Marzocca, *The Self-Represented Party—The Most Unique Party of Them All*, 59 ARIZ. 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." (citations omitted)). "In Arizona," moreover, "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." *Id.*
- 12 Data on file with the authors shows 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 e-filed (which is required by lawyers representing litigants).
- 13 Data on file with the authors shows that, from July 12, 2023, through November 30, 2023, of the 5,667 total initial eviction hearings in the Pima County Consolidated Justice Court, the average appearance rate was 62%, but that the appearance rate on Mondays was 60%, while the appearance rate on Tuesdays was 63%.
- 14 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*, L. Soc'y SCOTLAND (Aug. 14, 2023), <https://www.lawscot.org.uk/members/journal/issues/vol-68-issue-08/ai-opening-the-door-to-justice> (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.").
- 15 See Scott Rosenberg, *How We All Became AI's Brain Donors*, AXIOS (April 24, 2023), <https://www.axios.com/2023/04/24/ai-chatgpt-blogs-web-writing-training-data> (noting the use of internet data for the training of AIs).
- 16 Cf. Tatum Millet, *A Face in the Crowd: Facial Recognition Technology and the Value of Anonymity*, COLUM. J. TRANSNAT'L L. (Oct. 18, 2020), <https://www.jtl.columbia.edu/bulletin-blog/a-face-in-the-crowd-facial-recognition-technology-and-the-value-of-anonymity> ("Facial recognition technology . . . works by identifying unique details in peoples' faces, then comparing that facial data to other faces stored in a database, such as mugshot databases, DMV photos, and even social media.").
- 17 See Kushal Arora, Layla El Asri, Hareesh Bahuleyan & Jackie Chi Kit Cheung, *Why Exposure Bias Matters: An Imitation Learning Perspective of Error Accumulation in Language Generation*, 2022 FINDINGS ASS'N COMPUTATIONAL LINGUISTICS 700, 700 (2022) ("The distribution of these contexts seen during the generation phase might be very different from the ones encountered during the training phase. This mismatch is referred to as exposure bias.").
- 18 See Florian Schmidt, *Generalization in Generation: A Closer Look at Exposure Bias*, in PROCEEDINGS OF THE 3RD WORKSHOP ON NEURAL GENERATION AND TRANSLATION 157 (WNGT ed., 2019), available at <https://doi.org/10.3929/ethz-b-000393684>.
- 19 Lauren McCarthy, *A Wellness Chatbot Is Offline After Its 'Harmful' Focus on Weight Loss*, N.Y. TIMES (last updated June 9, 2023), <https://www.nytimes.com/2023/06/08/us/ai-chatbot-tesa-eating-disorders-association.html>.
- 20 See, e.g., Brianna Rauenzahn, Jamison Chung & Aaron Kaufman, *Facing Bias in Facial Recognition Technology*, REG. REV. (Mar. 20, 2021), <https://www.theregreview.org/2021/03/20/saturday-seminar-facing-bias-in-facial-recognition-technology>.
- 21 See Karen Weise & Cade Metz, *When A.I. Chatbots Hallucinate*, N.Y. TIMES (last updated May 9, 2023), <https://www.nytimes.com/2023/05/01/business/ai-chatbots-hallucination.html> ("[G]enerative A.I. . . . relies on a complex algorithm that analyzes the way humans put words together on the internet. It does not decide what is true and what is not. . . . The tech industry often refers to the inaccuracies as 'hallucinations.'").
- 22 See *id.*
- 23 See Larry Neumeister, *Lawyers Submitted Bogus Case Law Created by ChatGPT: A Judge Fined Them \$5,000*, ASSOC. PRESS (June 22, 2023), <https://apnews.com/article/artificial-intelligence-chatgpt-fake-case-lawyers-d6ae-9fa79d0542db9e1455397aef381c>.
- 24 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> ("The judge wrote in Thursday's sanctions order 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.'").
- 25 E.g., Hazelwood Sch. Dist. v. Kuhlmeier, 484 U.S. 260 (1988). Cf. 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").
- 26 See, e.g., *Armstrong v. Manzo*, 380 U.S. 545, 552 (1965); *Grannis v. Ordean*, 234 U.S. 385, 394 (1914).
- 27 See Judea Pearl & Dana Mackenzie, *AI Can't Reason Why*, WALL ST. J. (May 18, 2018), <https://www.wsj.com/articles/ai-cant-reason-why-1526657442> ("Put simply, today's machine-learning programs can't tell whether a crowing rooster makes the sun rise, or the other way around. . . . The questions "Why did this happen?" and "What if I had acted differently?" are . . . so far are missing from machines.").
- 28 See generally, e.g., Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218 (2019) (identifying some of the problems produced by biased training data).
- 29 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.
- 30 Kate Wells, *An Eating Disorders Chatbot Offered Dieting Advice, Raising Fears About AI in Health*, NPR (June 9, 2023), <https://www.npr.org/sections/health-shots/2023/06/08/1180838096/an-eating-disorders-chatbot-offered-dieting-advice-raising-fears-about-ai-in-hea>.
- 31 A very accessible overview appears at *Evaluation Methods of Justice Innovations*, STAN. LEGAL DESIGN LAB (last visited Nov. 30, 2023), <https://justiceinnovation.law.stanford.edu/resources/evaluation>.
- 32 See BUTLER, ET AL., *supra* note 2022, at 35–39.
- 33 See Marcia A. Testa & Donald C. Simonson, *The Use of Questionnaires and Surveys*, in CLINICAL AND TRANSLATIONAL SCIENCE: PRINCIPLES OF HUMAN RESEARCH 207, 207 (David Robertson & Gordon H. Williams eds., 2d ed. 2017) (2009) ("For many areas of clinical investigation, the interpretation of research results derived solely from clinical and laboratory data can be enhanced by information that reflects the patient's perspective of a disease condition or state of health.").
- 34 See generally Susan C. Stokes, *A Defense of Observational Research*, in FIELD EXPERIMENTS AND THEIR CRITICS: ESSAYS ON THE USES AND ABUSES OF EXPERIMENTATION IN THE SOCIAL SCIENCES 33 (Dawn Langan Teele ed., 2014) (discussing the role of confounding effects in observational research and potential tactics for mitigating them).
- 35 Gregg C. Fonarow, *Randomization—There Is No Substitute*, 1 JAMA CARDIOLOGY 633, 633 (2016) ("[R]andomization eliminates selection and other forms of bias, generates groups under study that are alike in all important aspects (except for the intervention received), and avoids confounding by measured and unmeasured confounding variables.").
- 36 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).