AI Chatbot

hello. i wrote this brief.

Is disclosure and certification of the use of generative AI really necessary?

BY MAURA R. GROSSMAN, PAUL W. GRIMM & DANIEL G. BROWN*

THE NEWS ABOUNDS WITH ARTICLES ON THE PROMISES — AND PERILS — OF GENERATIVE AI (GENAI) APPLICATIONS like ChatGPT, which create text or other content based on patterns learned from their training input. Depending on the writer's perspective, the future appears to be either utopian or dystopian in nature. But, as is usually the case, the truth falls somewhere in between: GenAI is a tool that has both benefits and risks. And regardless of one's viewpoint, the genie is already out of the bottle. GenAI applications are in widespread use, and billions of dollars are being invested in further development of this technology. The legal profession is not immune from these developments. Lawyers are already using GenAI for research and drafting purposes, and vendors are incorporating GenAI into eDiscovery tools as well. Its uses will only continue to proliferate.

Increasingly, judges are issuing individual standing orders that require litigants to disclose their use of GenAI and to submit certifications about their efforts to verify the accuracy of factual representations and case authority cited when using GenAI. Judges unquestionably have the inherent authority to issue orders and guidelines governing what parties can do in the cases pending before them, but little guidance has been offered on the use of GenAI in the justice system. While the impulse underlying the imposition of these standing orders is understandable — even commendable — real disadvantages can result. For example, some orders have been vague and ambiguous about the technologies they cover. Others have been overly broad — sweeping into their scope AI applications that do not produce final work product and that do not suffer from GenAI's propensity to "hallucinate" and generate erroneous output. Such orders can also infringe on attorney work product and may discourage the use of technology that might oth-

> While the impulse underlying the imposition of these standing orders is understandable — even commendable — real disadvantages can result.

erwise increase access to justice and reduce costs. And given the speed with which judges are issuing these orders, there has been a lack of consistency, which only adds to confusion and imposes additional burdens and costs on litigants who must — on pain of being sanctioned — make sure they know whether such an order governs and, if so, to adhere to it.

In this article, we outline what led to this judicial response, describe the various standing orders issued thus far, outline some of the concerns they raise, and discuss the technical issues and solutions currently available or on the horizon. Finally, we propose what we believe to be a better alternative: public notice and/or consistent, court-wide rules that are enacted following publication and public comment.

The Shot Heard 'Round the World: The Botched GenAI Filing

Alarms went off on May 27, 2023, when *The New York Times* reported that a court had issued an Order to Show Cause why plaintiff's counsel should not be sanctioned for papers they filed in opposition to a motion to dismiss[1] that were "replete with citations to non-existent cases."[2] The court asserted that "[s]ix of the submitted cases [in ▶

the opposition papers] appear[ed] to be bogus judicial decisions with bogus quotes and bogus internal citations."[3] It turned out that one of the attorneys in question had used ChatGPT to perform legal research, "a source that ha[d] revealed itself to be unreliable."[4]

In the immediate aftermath, several courts proactively issued standing orders to prevent such events in their own courtrooms. Just three days later, on May 30, 2023, Judge Brantley Starr of the U.S. District Court for the Northern District of Texas was the first to issue such a standing order.[5] He requires attorneys and pro se litigants appearing before him to file — on appearance in his court — a certificate indicating whether any portion of their filings would be drafted using GenAI tools. The standing order states in relevant part:

> All attorneys and pro se litigants appearing before the Court must, together with their notice of appearance, file on the docket a certificate attesting either that no portion of any filing will be drafted by generative artificial intelligence (such as ChatGPT, Harvey.AI, or Google Bard) or that any language drafted by generative artificial intelligence will be checked for accuracy, using print reporters or traditional legal databases, by a human being. . . . Any party believing a platform has the requisite accuracy and reliability for legal briefing may move for leave and explain why. Accordingly, the Court will strike any filing from a party who fails to file a certificate on the docket attesting that they have read the Court's judge-specific requirements and understand that they will be held responsible under Rule 11 for the

The solution proposed — a mosaic of inconsistent, individual standing orders — is not the best means to solve the problem, especially when existing rules can address the conduct at issue, and other institutions are better positioned to develop a more nuanced response.

contents of any filing that they sign and submit to the Court, regardless of whether generative artificial intelligence drafted any portion of that filing.

A week later, on June 6, 2023, Judge Michael M. Baylson of the U.S. District Court for the Eastern District of Pennsylvania issued an order requiring attorneys and pro se litigants to disclose the use of AI in drafting pleadings.[6] His order, however, was not limited to GenAI tools; rather, it referenced AI tools in general. His standing order stated:

> If any attorney for a party, or a pro se party has used Artificial Intelligence ("AI") in the preparation of any complaint, answer, motion, brief, or other paper filed with the Court, and assigned to Judge Michael M. Baylson, **MUST**, in a clear and plain factual statement, disclose that AI has been used in any way in the filing, and **CERTIFY**, that each and every citation to the law or the record in the paper, has been verified as accurate.[7]

Two days after that, on June 8, 2023, Magistrate Judge Gabriel A. Fuentes of the U.S. District Court for the Northern District of Illinois revised his standing order for civil cases,[8] to provide the following:

> The Court has adopted a new requirement in the fast-growing and fast-changing area of generative artificial intelligence ("AI") and its use in the practice of law. The requirement is as follows: Any party using any generative AI tool to conduct legal research or to draft documents for filing with the Court must disclose in the filing that AI was used, with the disclosure including the specific AI tool and the manner in which it was used. Further, Rule 11 of the Federal Rules of Civil Procedure continues to apply, and the Court will continue to construe all filings as a certification by the person signing the filed document and after a reasonable inquiry, of the matters set forth in the rule, including but not limited to those in Rule 11(b)(2). . . . Just as the Court did before the advent of AI as a tool for legal research and drafting, the Court will continue to presume that the Rule 11 certification is a representation by filers, as living, breathing, thinking human beings, that they

themselves have read and analyzed all cited authorities to ensure that such authorities exist and that the filings comply with Rule 11(b)(2).

On the same day, Judge Stephen Alexander Vaden of the U.S. Court of International Trade issued a standing order[9] not only requiring the disclosure of any GenAI program used for drafting but also requiring a representation that the use of such an application had not resulted in the disclosure of confidential or proprietary information to any unauthorized party. The relevant language of his order provides that:

Because generative artificial intelligence programs challenge the Court's ability to protect confidential and business proprietary information from access by unauthorized parties, it is hereby:

ORDERED that any submission in a case assigned to Judge Vaden that contains text drafted with the assistance of a generative artificial intelligence program on the basis of natural language prompts, including but not limited to ChatGPT and Google Bard, must be accompanied by:

(1) A disclosure that identifies the program used and the specific portions of test that have been so drafted;

(2) A certification that the use of such program has not resulted in the disclosure of any confidential business proprietary information to any unauthorized party; and it is further

ORDERED that, following the filing of such notice, any party may file with the Court any motion provided for by statute or the Rules of the Court of International Trade seeking any relief the

party believes the facts disclosed warrant.[10]

Not long thereafter, several Canadian courts followed suit. On June 23, 2023, the Court of King's Bench of Manitoba issued a *Practice Direction on the Use of Artificial Intelligence in Court Submissions*, advising that "when artificial intelligence has been used in the preparation of materials filed with the court, the materials must indicate how artificial intelligence was used."[11] Three days later, the Supreme Court of Yukon issued *Practice Direction General-29 on the Use of Artificial Intelligence Tools*,[12] which directed that "if any counsel or party relies on artificial intelligence (such as ChatGPT or any other artificial intelligence platform) for their legal research or submission in any matter and in any form before the Court, they must advise the Court of the tool used and for what purpose." Law360 Canada has reported that the Supreme Court of Canada "is among the courts mulling whether and what practice direction to issue to counsel and litigants about the use of artificial intelligence (AI) tools in the preparation of Supreme Court materials."[13]

Bringing Cannons to a Sword Fight: Are the Courts Overreacting?

We can certainly appreciate why courts throughout North America reacted swiftly and decisively to the GenAI mishap in the Southern District of New York — which, regrettably, was repeated again in a filing in the Tenth Court of Appeals in Waco, Texas, where an appellate brief contained "fabricated and non-existent citations."[14] No judge wants to discover that "[n]one of the three published cases cited actually exist in [a] [r]eporter," and that "[e]ach citation provide[d] the reader a jump-

cite into the body of a different case that ha[d] nothing to do with the proposition"[15] for which it was cited. But, we suggest that the solution proposed — a mosaic of inconsistent, individual standing orders — is not the best means to solve the problem, especially when existing rules can address the conduct at issue, and other institutions are better positioned to develop a more nuanced response.

We do not believe the courts that issued standing orders and practice directives intended to sow chaos or hamper innovation, but the result has been a lack of clarity and greater fear. Many different GenAI and other AI technologies exist, and some orders are not explicit about which technology use must be reported. For example, if a lawyer drafts a brief and uses Grammarly[16] to edit and revise their prose, does this need to be disclosed? Many online legal research databases already employ AI features for natural-language querying.[17] Must the use of such tools be reported, even though there is no risk of fake citations? And at what point does this reporting requirement begin to infringe on attorney work product and legal strategy?

Moreover, the landscape of potentially reportable GenAI applications is constantly changing. Most search engines[18] and word-processing systems,[19] for example, will soon embed the use of large language models (LLMs), a type of GenAI trained on massive data sets to recognize, translate, predict, or generate text in a human-like fashion. Rules of civil procedure should be technology-neutral and should not have to be revised with the introduction of each new technological development. No one can predict what the legal technology environment will look like two years from now, but ▶

the use of GenAI will almost surely be ubiquitous.

The legal profession is already sufficiently risk averse and technologically backward. These orders will impede innovation and chill the use of technology that could not only enable unrepresented parties to access the justice system, but also reduce the time and cost for those who can afford representation. We need a solution better tailored to the problem.

But first, it may be worth taking some time to better understand the problem itself. Below we briefly discuss the history of and technology underlying GenAI tools. The interested reader is referred to our forthcoming article, *The GPT Judge: Justice in a Generative AI World*,[20] for more detail.

Friend or Foe?: The Origins and Perils of GenAI

GenAI systems use deep-learning algorithms based on neural networks[21] to model written language, speech, music, or other pattern-based media. Typically, these systems are trained on vast collections of human-generated material — typically scraped from the internet — and then generate new work using the properties identified in the training dataset. GenAI systems can also be tuned to specific tasks. For example, one can fine-tune an AI model on the available artwork of a single artist and then generate thousands of new works in that style, potentially flooding the market with synthetic competition. Or the fine-tuning can be to a particular goal. One could, for example, train an LLM to write newspaper editorials from a particular political perspective. Some researchers and commercial entities have already developed special-purpose GenAI for conducting legal research or generating legal pleadings.[22]

The reason why ChatGPT 3.5 consistently correctly associates *Obergefell v. Hodges* with the topic of same-gender marriage is because the case is repeatedly mentioned in thousands of sentences about that subject in its training data, but citations to less well-known cases are less likely to be properly cited.

Recent technological advances have allowed much faster training of these models, as has the availability of larger training datasets, which explains what has appeared to be this technology's sudden emergence. In fact, ChatGPT, which incorporates OpenAI's GPT 3.5 model, is simply the latest in a series of generative pre-trained (GPT) LLMs that were introduced in early 2018.[23] Similarly, visual models like Dall-E 2, Midjourney, and Stable Diffusion are built upon previous models dating back to the early 2010s. Perhaps the primary reason for the recent emergence of so many such models is commercial: Corporations like Microsoft, Google,

OpenAI, and Meta are all trying to claim market dominance and have been rushing GenAI products to market.

From the perspective of the courts, the most important new developments in GenAI are LLM-based tools that, in response to a prompt, can generate text to fulfill the demands of that prompt. For example, a litigant might request that a tool "draft a complaint about a neighbor's noisy dog," or "find me a dog-noise case from Tennessee." Respectively, these tools will respond with a fully written complaint, or text that appears akin to a case citation.

Such instruments have the potential to exponentially expand efficiency and access to justice by reducing the time and expertise necessary to research and draft court filings. However, the goal of these LLMs is neither accuracy nor logical forms of argument per se, and they can be quite confident in presenting misinformation such that inaccuracies or fake case citations may nevertheless appear convincing. We address each challenge in turn.

How Does GenAI Sabotage the Truth?

One basic goal of GenAI is to model a style or a genre, like writing new poems in the style of Walt Whitman, or creating a satisfying werewolf romance story. These systems were not designed with accuracy as a goal, and they were not meant to engage in logical reasoning. Indeed, their primary purpose was to create new content. GPT methods sample from a probability distribution of relevant words and phrases, and while there may be some bias toward truthful results — to the extent the truth is more common among the sources from which GenAI draws — the model itself is unable to separate fact from fiction.[24] Newer LLMs attempt to create more

trustworthy content, but building in accurate citations and proper legal reasoning is a tall order.

This inability becomes especially problematic when one attempts to perform legal research using GenAI tools not built for that purpose. ChatGPT 3.5 routinely cites irrelevant or nonexistent cases, alongside relevant or real ones, because it is trying to fit the pattern of how one writes about the law; it is not necessarily trying to tell a true story. For example, in response to the prompt "find me a dog-noise case from Tennessee," ChatGPT 3.5 provided a response that claimed to be based on a Tennessee dog-noise case but actually miscited to a 2018 Texas Supreme Court medical-malpractice case (*Benge MD PLLC v. Williams*, Case No. 14-1057 (Tex. 2018)). And, when asked to write about the *Benge* case in the style of a newspaper article, ChatGPT 3.5 continued this incorrect pattern ("In a recent legal ruling, the Tennessee Court of Appeals addressed a contentious dispute between neighbors over incessant dog barking. The case of *Benge v. Williams* shed light on the complex issue of noise nuisances caused by pets and their potential impact on neighbors' quiet enjoyment of their property.").

The phenomenon at issue here, referred to as AI "hallucinations," is to be expected of LLMs; indeed, many consider it a feature rather than a bug. Recall that the training goal of LLMs is to emulate the textual style of the training dataset. Adding the word "not" or removing "only," for example, does not much change the overall fluency and apparent reasonableness of an LLM-generated sentence — but obviously can change the legal meaning dramatically. Similarly, a sentence in a GenAI-drafted legal brief may still fit the general structure of the text upon which the model was trained, regard-

less of whether the citations found in it are related to the subject. The reason why ChatGPT 3.5 consistently correctly associates *Obergefell v. Hodges*[25] with the topic of same-gender marriage is because the case is repeatedly mentioned in thousands of sentences about that subject in its training data, but citations to less well-known cases are less likely to be properly cited.[26] Even cases referenced in Wikipedia articles (such as Judge Grimm's *Mancia v. Mayflower Textile Services Co.* opinion),[27] can be misconstrued by ChatGPT 3.5. It claims that briefs citing that opinion focus on overtime pay and labor standards (the overall subject of *Mancia*), when, in fact, Judge Grimm's ruling focused on the parties' failure to cooperatively engage in the discovery process in violation of Fed. R. Civ. P. 26(g), the proposition for which the case has frequently been cited.

Newer or more purpose-built GenAI systems may eventually ameliorate this concern. For example, they can be trained to detect when a user is seeking a case citation and add a verification step to ensure valid and appropriate output. Also, as mentioned earlier, GenAI systems are now being built specifically for the purpose of legal research. For the time being, however, pro se filers will likely not have access — and may never have access — to the paid databases and specialized technologies used by lawyers and will instead turn to free, general-purpose GenAI systems (like ChatGPT 3.5).

Why Is GenAI so Good at Camouflage?

GenAI is hard to detect because its creators' primary goal was to develop a tool that would model the style of ordinary language, and because the models on which GenAI is based have quickly gotten better and massively more

complex. In particular, GenAI systems are now trained on larger and larger datasets — of largely unknown provenance — which include many different types of writing, languages, and levels of fluency. Training datasets typically include publicly available news sources, Wikipedia articles, government documents, Reddit posts, and much more. Since this training data includes many different styles of writing, the models learn the common and distinctive patterns of these various forms, and, on the surface, can convincingly mimic human-generated content.[28]

GenAI systems also make use of humans to identify when they create unconvincing (or unacceptable) outputs. This approach, Reinforcement Learning with Human Feedback (RLHF), allows the parameters of the model — a special kind of variable set during the training process — to be tuned so that it will create more believable (or acceptable) outcomes. A similar approach, Generative Adversarial Network (GAN), mimics a game between two AI participants. One GAN player generates new material, and then another attempts to discern what is fake and what is authentic by giving mathematical feedback to the generator, which updates and improves its output. This iterative process continues until the generator no longer improves. The better the distinguisher gets, the better the content generator gets, which explains why GenAI content can be hard to distinguish from human-generated content.[29] Some automated tools have sought to identify whether certain text is the output of an LLM or a human. LLMs often provide text that is more "unsurprising," in a mathematical sense, than text generated by humans (that is, the individual words in sentences are each, on average, more likely to occur in text written by

humans). This property can be used to detect AI-generated text.[30] However, in a recent experiment, one such tool incorrectly identified text written by nonnative English-speaking students (NNES) as having been crafted by GenAI. The smaller vocabularies and simpler sentence structure used by the NNES were flagged as hallmarks of AI generation.[31] Even OpenAI, the private company responsible for the creation of ChatGPT, recently withdrew its ChatGPT detection tool (GPTZero) for lack of accuracy.[32]

Other detection innovations have been suggested — for example, watermarking (i.e., hiding an invisible identifying marker in GenAI-produced text) that could allow one to later search for such an indicator in the text. However, since most LLMs do not watermark their output, one could simply use such an LLM as the last step in the creation process, asking the unmarked LLM to paraphrase the output of the watermarked LLM.[33] The fact is, those intent on mischief will always find ways to circumvent watermarks. Unfortunately, the arms race between content creators and detectors will continue, with no reason to believe that the typically less well-resourced content detectors will win.

Is There a Sufficient Arsenal of Weapons?: Existing Tools for Judges

Federal Rule of Civil Procedure 11
Part of our concern about the use of individual standing orders to regulate GenAI usage is that they impose on parties and litigants obligations that already apply under existing rules of civil practice and procedure and/or ethical obligations presently imposed on lawyers by state rules of professional responsibility. Most notably,

Rule 11 requires that all pleadings, motions, and other papers filed in civil cases be signed by a lawyer or, if the party is not represented by counsel, by the party themselves, and that signature carries with it certain assurances — assurances that render many of the recently AI-focused standing orders redundant.

Failure to sign a pleading obligates the court to strike the filing unless the omission is "promptly corrected after being called to the attorney's or party's attention."[34] The individual's signature on the pleading makes several specific representations to the court — namely, that "whether by signing, filing, submitting, or later advocating" what the pleading discusses, the "attorney or unrepresented party certifies that to the best of the person's knowledge, information, and belief, formed after an inquiry reasonable under the circumstances,"[35] that (1) it is not being presented for any improper purpose, like "to harass, cause unnecessary delay, or needlessly increase the cost of litigation"; (2) claims, defense, and legal contentions are "warranted by existing law or by a nonfrivolous argument for extending, modifying or reversing existing law or for establishing new law"; (3) "the factual contentions have evidentiary support or, if specifically so identified, will likely have evidentiary support after a reasonable opportunity for further investigation or discovery"; and (4) "the denials of factual contentions are warranted on the evidence, or if specifically so identified, are reasonably based on belief or a lack of information."[36]

Lawyers or pro se litigants who blindly rely on factual contentions taken from GenAI applications or who rely on — without independently confirming — cases cited by such applications clearly have failed to conduct a

reasonable inquiry, are filing a pleading that likely will cause unnecessary delay or increase litigation costs, are stating facts not based on existing law, and are presenting factual arguments without evidentiary support. The consequences of violating Rule 11 can be severe. A court may sanction any lawyer, law firm, or party that violated the rule or is responsible for it having been violated.

Thus, the standing orders described above appear to be redundant. If the consequences of failing to comply with Rule 11 do not adequately deter the conduct that courts have criticized regarding the use of GenAI, it is hard to imagine what additional deterrence a judge's individual standing order would lend.

Federal Rule of Civil Procedure 26(g)
On its face, Rule 11 applies only to pleadings, motions, and other "papers" and is inapplicable to discovery.[37] But this does not mean that there are no procedural impediments to a lawyer improperly using GenAI during the discovery phase of a civil case. Indeed, Rule 26(g)(1), which applies to "disclosures and discovery requests, responses, and objections," in civil cases also requires that every discovery-related disclosure, request, response, or objection must be signed by an attorney or party, if unrepresented.

As with Rule 11, the Rule 26(g)(1) signature "certifies that to the best of the person's knowledge, information, and belief formed after a reasonable inquiry" that: the disclosure is complete and correct as of the time it was made; and that a discovery request, response, or objection is (a) consistent with the discovery rules and warranted by existing law (or a nonfrivolous argument for extending, modifying, or reversing existing

law, or establishing new law), (b) is not interposed for an improper purpose (such as harassing an opponent, imposing unnecessary delay, or needlessly increasing the cost of litigation), and (c) is neither unreasonable nor unduly burdensome or expensive, considering the needs of the case, the amount in controversy in the case, and the importance of the issues at stake in the litigation.[38] If a party or attorney omits the required signature, the opposing counsel or party is under no duty to act on the discovery matter until it is signed, and the court must strike the unsigned discovery material unless the signature is promptly supplied when called to the attention of the lawyer or party. If a certification violates Rule 26(g), the offending lawyer and or party may be sanctioned.[39]

Accordingly, lawyers or parties who violate Rules 11 and 26(g) in connection with their use of GenAI in civil litigation are already subject to sanctions that can be strong medicine — depending on the extent of the violation — regardless of whether the presiding judge has issued their own standing order concerning the use of GenAI. Moreover, if widespread public humiliation over being sanctioned by a court for committing this kind of error is insufficient disincentive, the Rules of Professional Conduct also impose independent ethical obligations to refrain from the types of misconduct that have led courts to adopt standing orders prohibiting or regulating the use of GenAI applications.

American Bar Association Model Rules of Professional Conduct 1.1 Comment [8], 3.3, and 1.6 (and Their State-Law Equivalents)

All attorneys are required to be licensed by the states or provinces in

> **Lawyers or parties who violate Rules 11 and 26(g) in connection with their use of GenAI in civil litigation are already subject to sanctions that can be strong medicine — depending on the extent of the violation — regardless of whether the presiding judge has issued their own standing order concerning the use of GenAI.**

which they practice, and each jurisdiction has adopted rules of professional conduct that lawyers must follow, lest they be sanctioned or have their license suspended or revoked. In addition, almost all (at least in the U.S.) follow or are guided by the American Bar Association's (ABA's) Model Rules of Professional Conduct. Three of the Model Rules impose ethical duties relevant to the improper use of GenAI.

Model Rule 1.1 requires lawyers to provide clients with competent representation, which "requires . . . legal knowledge, skill, thoroughness, and

[reasonably necessary] preparation."[40] Comment [8] to Rule 1.1 provides that, "[t]o maintain the requisite knowledge and skill, a lawyer should keep abreast of changes in the law and its practice, including the benefits and risks associated with relevant technology."[41] GenAI is clearly a relevant technology to the practice of law today, and lawyers must understand its strengths and weaknesses to provide competent representation.

Model Rule 3.3 imposes an ethical obligation to demonstrate candor to courts and other tribunals and prohibits "mak[ing] a false statement of fact or law . . . or fail[ing] to correct a false statement of material fact or law previously made."[42] Citing nonexistent case law or misrepresenting the holdings of a case is making a false statement to a court. It does not matter if GenAI told you so.

Model Rule 1.6 prohibits lawyers from "reveal[ing] information relating to the representation of a client" without obtaining informed consent.[43] Entering confidential client information into a publicly available, third-party chatbot is inconsistent with this duty.

A lawyer who does not adequately understand the risks inherent in using GenAI, and who fails to independently verify the accuracy of factual matters and/or legal authority obtained from GenAI, has failed to represent their client competently. Moreover, a lawyer who uses factual information or legal authority obtained from GenAI in a pleading without independently confirming its accuracy fails to adhere to the obligation of candor to the court if those representations turn out to be false. Similarly, a lawyer who discloses information about the representation of a client to prompt a search using GenAI, without first having explained ▶

the risks and obtained consent, has failed to properly maintain confidentiality. To the extent the jurisdiction in question has adopted these model rules — or even the gist of them — none of these duties is likely to require a separate certification.

A judge who determines that a lawyer has used GenAI in a manner failing to conform with their ethical duties can refer the lawyer to the relevant licensing authority, and that body will likely initiate an ethics investigation that could result in sanctions, up to and including loss of their license to practice. Therefore, lawyers who engage in GenAI-associated misconduct risk more than the wrath of a single judge — they put their ability to practice law at risk. Bar associations and law societies should provide guidance and education to their members and remove this burden from individual judges.

Viewed both individually and collectively, existing rules of civil practice and procedure and ethical codes of conduct already provide adequate deterrence to the misuse of GenAI in litigation, and, if violated, provide sanctions that are at least as severe — if not more so — than can be imposed for failing to comply with a court's individual standing order.

### An Olive Branch: Public Notice and/or Local Rules

We believe that individualized standing orders are unnecessary, create unintended confusion, impose unnecessary burden and cost, and deter the legitimate use of GenAI applications that could increase productivity and access to justice. We do not, however, suggest that judges and courts should sit by idly and avoid engaging with issues regarding the use of GenAI in the justice system. Rather, if district

courts feel the need to address this issue, they can issue local rules that apply court-wide.[44] A well-crafted local rule governing the use of GenAI tools, adopted after publication and public comment, is more likely to address definitional and scope issues in a nuanced way and expose any unintended adverse consequences.

There is certainly no harm in individual judges including in their standing orders a *warning* to litigants about the risks inherent in using GenAI and the consequences of misrepresentations to the court. But, as mentioned earlier, sufficient deterrence may already exist. For the benefit of pro se litigants, in particular, courts can give notice to the public in general (e.g., on their websites) that the use of GenAI tools in connection with court filings must be consistent with the obligation to verify the accuracy of factual and legal representations, including validating all citations, and explain the potential sanctions for failure to do so. Additionally, we see no problem with requiring pro se litigants to disclose whether they have had any GenAI assistance in drafting their court filings. This would be similar to the mandates already imposed by certain state and local bar ethics committees that require either an attorney who has provided assistance to a party in drafting a court filing, but who has not entered an appearance as counsel for that party, to disclose to the court the assistance they provided, or for the pro se litigant to disclose that they received assistance in drafting the filing.[45]

It is evident that the use of AI applications — and GenAI in particular — will be increasingly common in the court system. However, we urge caution and restraint in imposing additional disclosure and certifica-

tion obligations — particularly when the scope of such requirements may be overbroad or ambiguous — which impose unnecessary and inconsistent burdens on litigants. It is possible, in this instance, that honey may work better than vinegar.
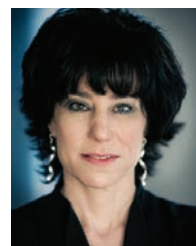
MAURA R. GROSSMAN, JD, PhD, is a research professor in the David R. Cheriton School of Computer Science at the University of Waterloo, an adjunct professor at Osgoode Hall Law School of York University, and an affiliate faculty member of the Vector Institute of Artificial Intelligence.

PAUL W. GRIMM is the director of the Bolch Judicial Institute and the David F. Levi Professor of the Practice of Law at Duke Law School. He is a retired district judge of the U.S. District Court for the District of Maryland, where he also served as a magistrate judge.

DANIEL G. BROWN, PhD, is a professor in the David R. Cheriton School of Computer Science at the University of Waterloo.

---

1 Benjamin Weiser, *Here's What Happens When Your Lawyer Uses ChatGPT*, N.Y. Times (May 27, 2023), https://www.nytimes.com/2023/05/27/

2   nyregion/avianca-airline-lawsuit-chatgpt.html.
Order to Show Cause, *Mata v. Avianca, Inc.*, No. 22-cv-1461 (PKC) (S.D.N.Y. May 4, 2023), 2023 WL 3696209 at *1.

3   *Id.*

4   Weiser, *supra*, note 1 (internal quotation removed).

5   *See*, Judge Specific Requirements, Mandatory Certification Regarding Generative Artificial Intelligence (Judge Brantley Starr, May 30, 2023), https://www.txnd.uscourts.gov/judge/judge-brantley-starr.

6   Standing Order Re: Artificial Intelligence ("AI") Cases Assigned to Judge Baylson (E.D. Pa. 2023), https://www.paed.uscourts.gov/documents/standord/Standing%20Order%20Re%20Artificial%20Intelligence%206.6.pdf.

7   *Id.* (emphases in original).

8   Standing Order for Civil Cases Before Magistrate Judge Fuentes (N.D. Ill. 2023), https://www.ilnd.uscourts.gov/_assets/_documents/_forms/_judges/Fuentes/Standing%20Order%20For%20Civil%20Cases%20Before%20Judge%20Fuentes%20rev%27d%205-31-23%20(002).pdf.

9   Stephen Alexander Vaden, J., Order on Artificial Intelligence (Ct. Int'l. Trade 2023), https://www.cit.uscourts.gov/sites/cit/files/Order%20on%20Artificial%20Intelligence.pdf.

10 *Id.* (emphases in original).

11 *See* https://www.manitobacourts.mb.ca/site/assets/files/2045/practice_direction_-_use_of_artificial_intelligence_in_court_submissions.pdf.

12 *See* https://www.yukoncourts.ca/sites/default/files/2023-06/GENERAL-29%20Use%20of%20AI.pdf.

13 Cristin Schmitz, *SCC Considers Possible Practice Direction on Use of AI in Top Court as More Trial Courts Weigh In*, Law360 Can. (July 7, 2023, 1:30 PM), https://www.law360.ca/articles/48377/scc-considers-possible-practice-direction-on-use-of-ai-in-top-court-as-more-trial-courts-weigh-in.

14 Lauren Berg, *Texas Appeals Court Calls out Seemingly AI-Generated Cites*, Law360 (July 26, 2023, 9:38 PM), available at https://www.law360.com/pulse/articles/1704217/texas-appeals-court-calls-out-seemingly-ai-generated-cites.

15 *Ex Parte Allen Michael Lee*, No. 10-22-00281-CR at 2 (10th Ct. App. TX July 19, 2023), 2023 WL 4624777 at *1.

16 *See* Grammarly, https://www.grammarly.com/ (last visited Aug. 21, 2023) ("generative AI writing assistant").

17 *See, e.g.*, *Westlaw Edge*, Thompson Reuters, https://legal.thomsonreuters.com/en/products/westlaw-edge (last visited Aug. 21, 2023) ("powered by AI-enhanced capabilities that can help you research more effectively and be more strategic").

18 *See, e.g.*, Will Knight, *Google Just Added Generative AI to Search*, WIRED (May 18, 2023 1:59 PM),

19 https://www.wired.com/story/google-io-just-added-generative-ai-to-search/.
See, e.g., Jared Spataro, *Introducing Microsoft 365 Copilot – Your Copilot for Work*, Official Microsoft Blog (Mar. 16, 2023), https://blogs.microsoft.com/blog/2023/03/16/introducing-microsoft-365-copilot-your-copilot-for-work/.

20 Maura R. Grossman, Paul W. Grimm, Daniel G. Brown, and Molly (Yiming) Xu, *The GPTJudge: Justice in a Generative AI World*, 23 Duke L. & Tech. Rev. (forthcoming Oct. 2023) (manuscript available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4460184).

21 Deep learning consists of a series of machine-learning algorithms made up of multiple layers: an input layer, one or more hidden layers, and an output layer. The method is referred to as "deep learning" because, unlike previous approaches, one layer can feed its output to the next layer. Each layer processes data in a manner inspired by the human brain, using interconnected nodes, hence the reason why they are often referred to as "neural networks."

22 *See, e.g.*, Casetext, https://casetext.com/ (last visited Aug. 21, 2023) ("Meet Co-Counsel – the world's first AI legal assistant."); Harvey.AI, https://www.harvey.ai/ (last visited Aug. 23, 2023) ("Unprecedented legal AI.").

23 Konstantinos I. Roumeliotis and Nikolaos D. Tselikas, *ChatGPT and Open-AI Models: A Preliminary Review*, 15 Future Internet, no. 6, 2023, at 192, https://doi.org/10.3390/fi15060192.

24 Mattew Hillier, *Why Does ChatGPT Generate Fake News References?*, TECHE (20 February, 2023), https://teche.mq.edu.au/2023/02/why-does-chatgpt-generate-fake-references/.

25 576 U.S. 644 (2015).

26 Another limitation of GenAI has to do with the date on which training of the system ceased. For example, ChatGPT 3.5 had a cutoff date of September 2021, so it cannot possibly cite to more recent cases. While newer, purpose-built tools will have more up-to-date cutoffs, unless the training is continually refreshed, there will always be issues involving (in)completeness.

27 253 F.R.D. 354 (D. Md. 2008).

28 AIContentfy Team, *AI Writing vs. Traditional Writing: Pros and Cons*, AIContentfy (July 10, 2023), https://aicontentfy.com/en/blog/ai-writing-vs-traditional-writing-pros-and-cons ("AI writing software is continuously improving by learning from vast amounts of existing written content, allowing it to generate increasingly accurate and contextually appropriate text.").

29 For further reading about GANs and RLHFs, *see generally* Jason Brownlee, *A Gentle Introduction to Generative Adversarial Networks (GANs)*, Machine Learning Mastery (July 19, 2019), https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/; Alex McFarland, *What is Reinforcement Learning*

from Human Feedback (RLHF), Unite.AI (March 29, 2023), https://www.unite.ai/what-is-reinforcement-learning-from-human-feedback-rlhf/; Zhang Ze Yu et al., *Fine-tuning Language Models with Generative Adversarial Feedback*, Cornell Univ. arXiv:2305.06176 [CS.CL], https://doi.org/10.48550/arXiv.2305.06176.

30 However, AI-to-human text converters can now take AI-generated text and add variety, uniqueness, and complexity to the content to bypass AI content detectors. *See, e.g.*, *AI to Human Text Converter*, Paraphrasing Tool, https://paraphrasingtool.ai/ai-content-bypass-tool/ (last visited Aug. 23, 2023).

31 Weixin Liang et al., *GPT Detectors Are Biased Against Non-Native English Writers*, 4 Patterns 1, 1–3 (2023), https://www.cell.com/patterns/fulltext/S2666-3899(23)00130-7#%20.

32 Benj Edwards, *Unsafe at Any Seed — OpenAI Discontinues its AI Writing Detector Due to "Low Rate of Accuracy,"* Ars Technica (July 26, 2023, 3:51 PM), https://arstechnica.com/information-technology/2023/07/openai-discontinues-its-ai-writing-detector-due-to-low-rate-of-accuracy/.

33 Zygmunt Zając, *SpamGPT: Watermarking Large Language Models*, FastML (March 14, 2023), https://fasml.com/spamgpt-watermarking-large-language-models/.

34 Fed. R. Civ. P. 11(a).

35 Red. R. Civ. P. 11(b).

36 Fed. R. Civ. P. 11(b) (1)–(4).

37 Fed. R. Civ. P. 11(a), (d) "[Rule 11] does not apply to disclosures and discovery requests, responses, objections, and motions under Rules 26 through 37 [which deal with discovery]."

38 Fed. R. Civ. P. 26(g)(1).

39 Fed. R. Civ. P. 26(g)(3).

40 Model Code of Prof. Conduct r. 1.1 (Am. Bar Ass'n 2020).

41 Model Code of Prof. Conduct r. 1.1 cmt. [8] (Am. Bar Ass'n 2020).

42 Model Code of Prof. Conduct r. 3.3(a)(1) (Am. Bar Ass'n 2020).

43 Model Code of Prof. Conduct r. 1.6(a) (Am. Bar Ass'n 2020).

44 *See* 28 U.S.C. § 2071(b) ("Any rule prescribed by a court, other than the Supreme Court, under subsection (a) shall be prescribed only after giving appropriate public notice and an opportunity for comment."); Fed. R. Civ. P. 83(a)(1) ("After giving public notice and an opportunity for comment, a district court, acting by a majority of its district judges, may adopt and amend rules governing its practice.").

45 *See* ABA Comm. on Ethics & Pro. Resp., Formal Op. 07-446, at 1–2, nn.4–5 (2007) (finding no unethical conduct for lawyers providing legal assistance to pro se litigants without disclosing the nature of their assistance).