By Hon. Cynthia A. Norton and Prof. Nancy B. Rapoport[1]

I'll Be the Judge of That!

Hon. Cynthia A. Norton
U.S. Bankruptcy Court (W.D. Mo.) Kansas City

Prof. Nancy B. Rapoport
UNLV William S. Boyd School of Law; Las Vegas

Hon. Cynthia Norton is Chief Judge of the U.S. Bankruptcy Court for the Western District of Missouri. She has been on the bench since 2013. Nancy Rapoport is a University of Nevada Las Vegas distinguished professor and the Garman Turner Gordon professor at the William S. Boyd School of Law.

Jason Willick's recent *Washington Post* opinion piece[2] alerted us to a study by Profs. Eric Posner and Shivan Saran[3] that sought to determine whether the judgment of artificial intelligence (AI) is superior to that of real-life federal judges. Because one of us actually is a federal judge and the other one used to be a bit of a statistics geek, we dove right in and read the Posner/Saran paper.

Much like the LawGeex study that showed that generative AI (GenAI) was better than sophisticated lawyers when it came to reviewing language in NDAs,[4] the Posner/Saran paper demonstrated that when LLMs and human judges are given the same packet of information (a set of facts, briefs, a summarized lower court opinion and a summarized statute), OpenAI's GPT-4o stuck closer to precedent than the human judges.[5] In the study, the facts were presented in two ways: one that made the defendant more sympathetic, and one that did not. GPT-4o appears to have been developed in a way that — try as the study's authors might — refuses to factor in "soft" factors in its decision-making.

As Willick summarized the Posner/Saran paper: "The bottom line is that real judges appear to be more easily swayed by 'legally irrelevant' factors than [AI] presented with the same material. That result, however, contains a delicious duality: It highlights either the manifest fallibility of human judges or their superior wisdom. Perhaps they are one and the same."[6] If you get a chance, read the Posner/Saran study, which is much richer than we are describing in this short paragraph.[7]

However, the study — and the *Washington Post* write-up — got us thinking: Just how do judges think about judging, and are there ways in which AI-based judging can be better than people whom the Senate has confirmed (or a circuit or district court has appointed)? We spent some time visiting these issues.

Prof. Rapoport: Judge Norton, when you went to "baby judges' school," what were you taught about judicial decision-making? What were the basic "dos" and "don'ts"?

Chief Judge Norton: That question alone could be the subject of another article; should I ask ChatGPT to write it for me? Seriously, though, we learn some basics about how the brain works (heuristics and slow and fast thinking), how to be mindful of external factors that may impact decision-making (like being hungry or tired), how to make parties feel heard, how to avoid making decisions based on our own implicit biases (or blind spots, as I like to think about it), and what considerations are important in deciding whether to write or rule from the bench, among other things.

Prof. Rapoport: I know that every case is different — different facts, different precedents, different lawyers, even — but when you're wrestling with a decision, do you have a typical way that you approach weighing the evidence?

Chief Judge Norton: First, I try to be as prepared as I can be. I will have read the briefs, done independent research and discussed the issues with my law clerks. Most importantly, I try to keep an open mind; I've been surprised at how many times I have anticipated I would rule one way, but after hearing the evidence and arguments I've ruled the other way. I'm also cognizant that as bankruptcy judges, we have some limited ability to "do equity" when we can.

Prof. Rapoport: It seems that GenAI is pretty good at applying legal precedents to new facts (at least as good as law students are). Part of a judge's role is to weigh the factors in an individual case, which is way more nuanced that what GenAI can do (yet). How do you weigh things

1 Judge Norton also is an advisory board member of ABI's Midwestern Bankruptcy Institute. Prof. Rapoport is an affiliate professor of business law and ethics in the Lee Business School at UNLV, and a past ABI Vice President-Research Grants.

2 Jason Willick, "A Provocative Experiment Pits AI Against Federal Judges," *Wash. Post* (March 17, 2025), washingtonpost.com/opinions/2025/03/17/artificial-intelligence-federal-judges-experiment (unless otherwise specified, all links in this article were last visited on June 23, 2025).

3 Eric A. Posner & Shivan Saran, "Judge AI: Assessing Large Language Models in Judicial Decision-Making," Coase-Sandor Inst. for Law and Econs. Research Paper No. 25-03, papers.ssrn.com/sol3/papers.cfm?abstractid=5098708.

4 Michael Simon, Alvin F. Lindsay, Loly Sosa & Paige Comparato, "*Lola v. Skadden* and the Automation of the Legal Profession," 20 *Yale J.L. & Tech.* 234, 282 (2018). The LawGeex study found that GenAI was both more accurate (94 percent accuracy vs. 85 percent accuracy) and faster (92 minutes vs. 26 seconds). *Id. But see* Mariya Yao, "Chihuahua or Muffin? My Search for the Best Computer Vision API," FreeCodeCamp (Oct. 12, 2017), freecodecamp.org/news/chihuahua-or-muffin-my-search-for-the-best-computer-vision-api-cbda4d6b425d (determining whether computer or human eye is better at distinguishing pictures of chihuahuas from pictures of blueberry muffins).

5 Posner & Saran, *supra* n.3, at 13 ("In short, GPT [is] more formalist (in the sense of following precedent), and judges are less so (in the sense of being swayed by sympathy).").

6 Willick, *supra* n.2.

7 It's OK to skip over the statistics part and just read the rest. But here's one of the study's main points: "[T]he apparent weakness of human judges is actually a strength. Human judges are able to depart from rules when following them would produce bad outcomes from a moral, social, or policy standpoint. Human judges also vary in their judicial philosophies and decision-making strategies." Posner & Saran, *supra* n.3, at 28. The study also cracked us up when describing why it's more likely that GPT fell short: It tended to make the same decisions that law students did. "[O]ne clue we have that GPT performed less well than human judges is [that GPT] performed nearly the same as students. Unless we think that students are better judges than professionals, we are forced to conclude that GPT is a worse judge than the professionals."). *Id.*

like how a decision in one case might create a bad precedent in future cases?

Chief Judge Norton: I'm bound to follow the law, so if the applicable law and the evidence before me are going to compel me to create a bad precedent, I will hold my nose and do it. Now, I have some leeway in how I issue the decision; I will likely rule from the bench, rather than issue a written opinion, or I may drop a footnote to explain that although I don't like the result, I'm bound by, for example, the circuit-level authority on student loan discharges. But I might urge the circuit to reconsider its approach. I also tend to give lots of warning shots over the bow to lawyers that they may get a result that comes back to bite them and suggest that they settle. Lawyers don't always heed those warning shots, however.

Prof. Rapoport: What should we do about the "bad facts make bad law" problem? It seems as though GenAI isn't at all concerned with that type of situation?

Chief Judge Norton: Apropos of the warning shots I like to give to lawyers, all I can say is that you can lead a horse to water but you can't make it drink. I'd bet a GenAI-generated decision wouldn't incorporate humor or a cheesy old saying into its answer, would it, Nancy?

Prof. Rapoport: Well, I've discovered that GPT's DALL-E can be sarcastic, albeit inadvertently.[8] But you're right: Computers don't really have a sense of humor. Still, I'll bet that there is some scutwork that GenAI might be able to do to make your life easier, such as pull evidence from a really dense record so that you can cite to that evidence in your opinions.[9] Have you experimented with that?

Chief Judge Norton: Although some judges jumped in early and are reporting that using GenAI helps them save time in summarizing transcripts, creating timelines or even brainstorming basic background about an unfamiliar area of law, many of us have been reluctant to use something like ChatGPT on our government devices, at least until we were sure we were aware of the risks. We are just now getting access to licenses for various programs, so I am anxious to start using it, and I have received my first training.

Prof. Rapoport: In *Snell v. United Specialty Ins. Co.*,[10] Judge Kevin Newsom wrote a concurring opinion in which he used GenAI to help him determine whether something fit within the definition of "landscaping." That was an intriguing use of AI, and he was transparent in his use of it, but what risks are there if a judge goes outside of the record by using AI?

Chief Judge Norton: The important thing to note about Judge Newsome's use of GenAI was that what he learned did not impact or decide the issue on which the appeal was decided. As the ABA Working Group[11] emphasizes, an independent, competent, impartial and ethical judiciary is indispensable to justice in our society. So, in using GenAI, a judge must be aware of the risks that the results may not be accurate; that the judge might be tempted to rely on extrajudicial information and influences that parties do not have the opportunity to address or rebut; and that the quality of the output depends on the quality of the prompt. As the Working Group notes, overreliance on GenAI may undermine the essential human judgment that lies at the heart of judicial decision-making.

Prof. Rapoport: What if lawyers use GenAI as part of their argument — not just in doing first drafts (and there are now several courts that have local rules about lawyers' AI use[12]), but by demonstrating, in essence, "what is known" about a concept?

Chief Judge Norton: I know other judges in good faith disagree, but I personally don't think courts should be in the business of discouraging the use of a technology that may save lawyers time and money, so long as the lawyers using GenAI verify the results, comply with applicable ethics rules in their jurisdiction regarding the use of GenAI, and comply with any applicable court orders requiring disclosure of the use of GenAI in documents filed with the court.

Prof. Rapoport: Any other "life lessons" on this topic for judges considering the use of GenAI in their chambers?

Chief Judge Norton: I remember thinking it was the end of the world to switch from paper to electronic filing and from WordPerfect to Word. Of course, we adapted and are better for it. All new technologies have the promise to help us do our jobs better and more efficiently. But as one judge so eloquently put it, "much like a chain saw or other useful [but] potentially dangerous tools, one must understand the tools they are using and use those tools with caution. It should go without saying that any use of artificial intelligence must be consistent with counsel's ethical and professional obligations. In other words, the use of artificial intelligence must be accompanied by the application of actual intelligence in its execution."[13] As of the date of this writing, there have been dozens of lawyers sanctioned for citing hallucinated cases.[14] Please don't be the lawyer that makes me use my chainsaw on you! abi

8 *See* "Maybe Dall-E3 Has a Sense of Humor. Or Maybe It's Sadistic.," *Nancy Rapoport's Blog* (Jan. 27, 2025), nancyrapoports.blog/2025/01/27/maybe-dall-e3-has-a-sense-of-humor-or-maybe-its-sadistic.

9 *Cf.*, "How Arbitrators Are Harnessing Artificial Intelligence," Am. Arbitration Ass'n (Feb. 20, 2024), adr.org/blog/how-arbitrators-are-harnessing-artificial-intelligence.

10 *Snell v. United Specialty Ins. Co.*, 102 F.4th 1208 (11th Cir. 2024).

11 "Navigating AI in the Judiciary: New Guidelines for Judges and Their Chambers," 26 *The Sedona Conference Journal* (forthcoming 2025).

12 This is part of the frustration that we both share in reading local rules that don't define which type of AI the rules cover.

13 *Mid Central Operating Engineers Health and Welfare Fund v. HoosierVac LLC*, No. 2:24-cv-00326, 2025 WL 574234, at *2 (S.D. Ind. Feb.21, 2025) (imposing $15,000 sanction and disciplinary referral upon lawyer who cited nonexistent cases generated by GenAI).

14 *See, e.g.*, Cecily Mauran, 120 Court Cases Have Been Caught with AI Hallucinations, According to New Database," *Mashable* (May 27, 2025), mashable.com/article/over-120-court-cases-caught-ai-hallucinations-new-database.