

Circuit Judge Writes An Opinion, AI Helps: What Now?

By **David Zaslowsky** (June 4, 2024)

Articles about artificial intelligence appear daily in many legal publications — including this one. But, on May 28, U.S. Circuit Judge Kevin Newsom of the U.S. Court of Appeals for the Eleventh Circuit wrote a concurring opinion in *Snell v. United Specialty Insurance Co.* for the specific purpose of floating an idea about how judges might use generative AI.



David Zaslowsky

Judge Newsom undertook this potential "heresy" — to use his word — with a light touch, as indicated in the following excerpt from the beginning of the opinion:

Here's the proposal, which I suspect many will reflexively condemn as heresy, but which I promise to unpack if given the chance: Those, like me, who believe that "ordinary meaning" is the foundational rule for the evaluation of legal texts should consider — consider — whether and how AI-powered large language models like OpenAI's ChatGPT, Google's Gemini, and Anthropic's Claude might — might — inform the interpretive analysis. There, having thought the unthinkable, I've said the unsayable. Now let me explain myself.[1]

I touch below on the highlights of Judge Newsom's interesting dissertation, but the entire opinion should be reviewed to understand the judge's full journey.

The underlying lawsuit was brought by Snell, a landscaper, who sought coverage under an insurance policy in connection with an accident that occurred during his installation of a ground-level trampoline in a client's backyard. The insurance company declined coverage on the ground that the accident did not arise from "landscaping." It won summary judgment.

The Eleventh Circuit affirmed. Its decision was ultimately based on the facts that (1) Snell's insurance application denied that his work included "any recreational or playground equipment construction or erection," and (2) Alabama law makes the application part of the policy.

As Judge Newsom explained, however, because that outcome was not always the apparent one, he spent many hours thinking about the "ordinary meaning" of the word "landscaping." He started with dictionaries, as any textualist would. But, he wrote:

[I]t was midway along that journey that I had the disconcerting thought that underlies this separate writing: Is it absurd to think that ChatGPT might be able to shed some light on what the term "landscaping" means? Initially, I answered my own question in the affirmative: Yes, Kevin, that is positively absurd. But the longer and more deeply I considered it, the less absurd it seemed.

He ultimately submitted two prompts to ChatGPT: "What is the ordinary meaning of 'landscaping'?" and "Is installing an in-ground trampoline 'landscaping'?" He also submitted the same prompts to Google's Bard (since replaced by Gemini). The responses supported the conclusion that "landscaping" could include Snell's work, which was consistent with the judge's personal ordinary-meaning understanding of the term.

Judge Newsom then discussed the advantages of using generative AI as a tool in the toolkit to inform an ordinary-meaning analysis.

First, the foundation of the ordinary-meaning rule is the common speech of common people. Large language models, or LLMs, that undergird generative AI systems like ChatGPT are literally "taught" using data that aim to reflect and capture how individuals use language in their everyday lives.

Second, the more powerful models are capable of understanding context, making them "high-octane language-prediction machines capable of probabilistically mapping, among other things, how ordinary people use words and phrases in context."

Third, LLMs are readily accessible to judges, lawyers and ordinary citizens. And, fourth, LLMs are arguably more transparent than dictionaries in that, while the latter are ordinarily taken for granted, their construction is not always self-evident in terms of who compiles them, and by what criteria.

Judge Newsom then proceeded to discuss the potential pitfalls.

The most significant is that LLMs can "hallucinate" — provide text or results that an AI program presents as fact, but which is, in fact, false. Second, LLMs do not capture offline speech, and thus might not fully account for underrepresented populations' usages — a criticism that falls under the category of "bias."

Third, "there's a risk that lawyers and judges might try to use LLMs strategically to reverse-engineer a preferred answer — say, by shopping around among the available models or manipulating queries" but, as he noted, that is an evergreen issue.

Judge Newsom's bottom line is as follows:

My only proposal — and, again, I think it's a pretty modest one — is that we consider whether LLMs might provide additional datapoints to be used alongside dictionaries, canons, and syntactical context in the assessment of terms' ordinary meaning. That's all; that's it.

This proposal is indeed "modest," perhaps to a fault. But Judge Newsom's opinion is nevertheless significant because it offers a circuit court judge discussing in a thorough and meaningful way a real use case for LLMs in the adjudicative process.

Hopefully, this will be just a jumping-off point. Judge Newsom makes absolutely clear that he is "not, not, not" suggesting that AI be used for rendering judgment. But why should that not be a near, short-term goal — at least for some types of cases?

Online dispute resolution that relies on AI is a reality in private industry. Ebay, for example, is one of the largest and most well-known users of ODR.

Its four-step dispute resolution process involves two steps that are completely automated, followed, if necessary, by mediation and then arbitration. The algorithm that tries to find a mediated settlement is trained on the millions of disputes that have arisen at eBay. More than 90% of disputes are settled with no human interaction, and eBay reports very high user satisfaction rates.[2]

The judicial system should be encouraged to work with the providers of ODR programs to

develop and pilot AI systems that can be used to resolve high-volume but relatively low-stakes disputes that typically have relatively discreet issues — which, for example, probably describes a good number of, though certainly not all, debt collection lawsuits.

Parties in such cases are, in any event, often unrepresented by lawyers. Furthermore, there is an enormous burden on the judicial system for each such case to be processed and ultimately determined by a human judge. Development of a working AI system that would allow parties the option of having such cases mediated in an automated way would result in cases being resolved in a much more efficient and expedient manner.

Such systems, at least initially, should not be mandatory processes for deciding cases, which in any event might implicate Seventh Amendment issues, but an option for parties to opt into. Positive user feedback and demonstrable benefits of such an AI system should be relied on as the drivers for participation. Indeed, parties should even be given the right to reject the suggested resolution of the AI mediator and to opt back into court adjudication.

While the initial reaction to such a suggestion might be that the losing side would always opt back into court adjudication and thus defeat the hoped-for efficiency gains, eBay's results do not support that conclusion. A likely explanation is that an optimal mediated settlement does not leave either side believing that it lost. Rather, it would appear that a determinative factor is whether a party believes that it was treated fairly.

In addition, a party would have to weigh the possibility that the court adjudication that she would be opting back into would take even more time and effort, but could very well result in a similar outcome. If these types of AI systems were successful, it could help alleviate the court congestion problem that plagues so many court systems, especially state court systems.

Another interesting aspect of Judge Newsom's decision is that it highlights the lack of any guidance in the U.S. on how courts should — or, perhaps it is better to say, "may" — use generative AI. Judge Newsom is refreshingly candid about his journey toward using AI. But is it being unduly cynical to suggest that there are probably other judges, or law clerks, who have used generative AI without being as transparent as Judge Newsom?

Is one going out on a limb in suggesting that judges have used it, or may soon use it, as a primary basis for decision making? Might law clerks, who, speaking in general terms, are probably more facile with this technology, be using it?

These are not far-fetched suggestions. Judge Newsom moved almost naturally from asking a question about the meaning of "landscaping" to asking whether "installing an in-ground trampoline [is] 'landscaping'" — the ultimate issue in the case.

The Guardian reported that a judge in Colombia admitted to using ChatGPT to help decide whether an autistic child's insurance should cover all the costs of his medical treatment.[3] According to the article, the judge "asked ChatGPT the precise legal matter at hand: 'Is an autistic minor exonerated from paying fees for their therapies?'" The response, that minors were exempt under Colombian law, was the decision rendered by the court.

In 2023, Sky News reported that Lord Justice Birss, a Court of Appeal judge in England who specializes in intellectual property law, admitted that he used "jolly useful" Chat GPT in writing an opinion.[4] Speaking at a conference in September 2023, he said:

I asked ChatGPT can you give me a summary of this area of law, and it gave me a paragraph. I know what the answer is because I was about to write a paragraph that said that. But it did it for me and I put it in my judgment.

To his credit, the judge also warned against relying on AI for topics on which the person does not know anything.

Furthermore, in contrast to the U.S., on Dec. 12, 2023, the Courts and Tribunals Judiciary issued guidance regarding AI for judges in England and Wales.^[5] The guidance warned about hallucinations, saying that "the information provided may be inaccurate, incomplete, misleading, or biased."

It also mentioned the importance of not entering confidential information into a public AI chatbot and warned against biases inherent in many LLMs because of the way they are trained. The guidance identified tasks such as summarizing large bodies of text and writing presentations as potentially useful ones in which AI can be used. And the guidance said that legal research in areas unknown to the user, as well as legal analysis, were the types of tasks that were not recommended.

In my opinion, though, one of the most interesting parts of the guidance is the following statement: "Judges are not generally obliged to describe the research or preparatory work which may have been done in order to produce a judgment. Provided these guidelines are appropriately followed, there is no reason why generative AI could not be a potentially useful secondary tool."

The most reasonable reading of this statement is that judges do not have to disclose their use of AI. To the extent that the recommendations in the guidance are followed and generative AI tools are (1) used only as a "secondary" tool, and (2) not used for researching unknown areas or legal reasoning, it is understandable why disclosure would not be required.

However, it seems inevitable that judges will come to rely on generative AI tools in making decisions — as the Colombian judge has done already. The tools to be used for such purposes will likely, and hopefully, be ones developed specifically for the legal industry because lawyers and judges should take very seriously the admonition given by the more well-known AI platforms that they should not be relied on for legal research.

In any event, when that threshold is passed, there probably should be a requirement that this type of use needs to be disclosed, the precise nature of which will have to be considered by those responsible for drafting judicial ethical rules. Given the seriousness of the risks in using generative AI — hallucinations and bias being the paramount ones — there should be transparency on judges' use of such tools when they play a primary role in decision making, thereby creating an environment in which there are checks in place and mechanisms for identifying errors.

Judge Newsom's concurring opinion is enlightening. But it is just step one down the path of, hopefully, developing a sound and coherent basis for LLMs and generative AI to be used by the judiciary. At the same time, there is the reality that the technology is developing at exponential rates, which far outstrips the pace at which the judiciary usually moves. Whether and how those two realities can coexist will be interesting to watch.

David Zaslowsky is a partner at Baker McKenzie.

The opinions expressed are those of the author(s) and do not necessarily reflect the views of their employer, its clients, or Portfolio Media Inc., or any of its or their respective affiliates. This article is for general information purposes and is not intended to be and should not be taken as legal advice.

[1] Snell v. United Specialty Insurance Co., Case no. 22-12581 (11th Cir. May 28, 2024) (concurring opinion).

[2] See Barton, Benjamin, Rebooting Justice: ODR is Disrupting the Judicial System, Law Practice July/August 2018 (available at Rebooting Justice: ODR is Disrupting the Judicial System (utk.edu), last accessed May 31, 2024).

[3] <https://www.theguardian.com/technology/2023/feb/03/colombia-judge-chatgpt-ruling> (last accessed May 31, 2024).

[4] <https://news.sky.com/story/british-judge-admits-using-jolly-useful-chatgpt-to-write-ruling-12961647> (last accessed May 31, 2024).

[5] <https://www.judiciary.uk/wp-content/uploads/2023/12/AI-Judicial-Guidance.pdf> (last accessed May 31, 2024).