Drawing the Line: AI, Evidence, and Advocacy

Webinar Transcript | November 6, 2026

Sponsored by the Duke Center on Law & Technology
and the Bolch Judicial Institute at Duke Law

Participants

- **Kim Mueller** (Moderator): David F. Levi Professor of the Practice of Law and Director, Bolch Judicial Institute, Duke Law School
- **Judge Marty B. McGee** (MJS '25): North Carolina Superior Court Judge; Inaugural RAILS Fellow
- **Aparna Komarla**: Data scientist; Founder, Redo.io; Inaugural RAILS Fellow
- **Kelli Raker**: Duke Center on Law & Tech, Duke Law School (welcome remarks and audience Q&A)

This transcript has been lightly edited for readability: filler words, false starts, and repetitions have been removed, and names and terms have been corrected. Clarifications are added as needed using [] brackets. Timestamps refer to the webinar recording.

Opening Remarks

Kelli Raker [00:00:25]: Hello, everyone. Thank you so much for joining us today.

Kelli Raker [00:00:32]: My name is Kelli Raker, and I'm with Duke Law School. I'm thrilled to welcome you to our webinar, "Drawing the Line: AI, Evidence, and Advocacy," sponsored by the Duke Center on Law & Technology and the Bolch Judicial Institute at Duke Law. Over the next hour, you'll hear from two of our inaugural Responsible AI in Legal Services, or RAILS, Fellows about what they've been working on over the last year. When I hop off the screen, I'll drop a link in the chat to the page on our website about our fellows, which links to their publications. It has truly been my pleasure to work with all three of our fellows on their projects over the last year, and I'm really looking forward to today's conversation.

I have three quick housekeeping items. First, we are recording the webinar so that we can send it to you next week in case you have to leave early or want to share it with a colleague. Second, we are using the Q&A function today. If you'd like to submit a question to any of our speakers, we'll do our best to address a few of those at the end of the program. And last, if you have any

technical issues, please use the chat to message me and let me know if there's any trouble connecting, with audio, or anything like that.

Thank you again so much for being here. I'm going to go ahead and turn the program over to the director of the Bolch Judicial Institute, Kim Mueller, and I invite our speakers to turn on their cameras for the conversation. Thank you all, and enjoy.

Welcome and Introductions

Kim Mueller [00:02:09]: Thank you so much, Kelli. It is my pleasure to be here. I am Kim Mueller, the director of the Bolch Judicial Institute at Duke Law School. I'm also a professor of the practice of law—but not the practice of AI in the law, so I am a student here today as well. I am a former federal trial judge. I served on the U.S. District Court for the Eastern District of California as a district judge for fifteen years, and as a magistrate judge for eight years before that. I retired before I really had to think too deeply about AI. I am a Duke graduate myself: the Master of Judicial Studies program, which Judge McGee and I participated in together, was a great gift, and we did start to learn about AI there.

It is the Bolch Judicial Institute's pleasure to co-host today's really important discussion. We'll move through several sets of questions, and we hope to have at least five to ten minutes at the end for your questions, so please let us know what you'd like us to address. I'd now like to ask Judge McGee and Ms. Komarla to introduce themselves and, after you tell us who you are, to share a few thoughts about the value of the RAILS Fellowship to you and your work. Judge McGee?

Judge Marty B. McGee [00:03:25]: Well, thank you very much. I appreciate the opportunity to be here with you all today. I am Marty McGee, a North Carolina Superior Court judge. I've been a trial judge for twenty-six years, at the district court level and at the superior court level. I had the pleasure of participating in the Duke LLM (judicial studies) program with Judge Mueller, which was a life-changing experience for me. I was fortunate enough to be selected as one of the RAILS Fellows, and it has given me the opportunity to use AI-specific tools—Harvey, CoCounsel, and Protégé—in my legal work.

My goal was to try to produce a proposed model order, to move from the scattered approach we have now toward central organizing principles that we could apply nationwide. I simply could not have done that without the access to the tools that the RAILS program provided. Judges have very little opportunity to use these tools, and I think expanding that access would be incredibly helpful. I so much appreciate the opportunity to have used those tools, and getting to work with Professor Jeff Ward and Kelli Raker throughout this process has been a wonderful experience.

Kim Mueller [00:04:51]: Thank you, Judge McGee. Ms. Komarla?

Aparna Komarla [00:04:55]: Hi, everyone. Thanks so much, Kim and Kelli, for organizing this program. I'm Aparna. I'm a data scientist and a technologist, and I'm the founder of Redo.io, a legal AI platform that supports resentencing in California. As of last month, I'm also a graduate student in the Symbolic Systems program at Stanford, where I'm studying computer science and math. I've really enjoyed this fellowship so far. I'm very grateful for all of the resources we've been able to access through Harvey and other legal research tools, as Marty mentioned. It's also been a great environment to connect with folks who are practicing law from many different angles and perspectives. It's been a great experience so far.

The Function Rule: AI, Advocacy, and Evidence

Kim Mueller [00:05:48]: Thank you, each of you, and again, thank you for being here. Let's move into our first set of questions, where we'll explore AI when it comes to evidence and, particularly for Judge McGee, disclosure issues. Judge McGee, you've come up with what I think you call the Function Rule. You draw a bright line in your framework between AI used to draft a brief and AI that materially affects evidence. Can you walk us through that framework and help us understand the moment that distinction crystallized for you?

Judge Marty B. McGee [00:06:29]: Yes, certainly. When we started our LLM program back in May of 2023, ChatGPT had been released just five months earlier, so AI was certainly new to me, and there was a lot of focus on hallucinations. We had the opportunity to be exposed to excellent professors—Judge Paul Grimm, Maura Grossman, and Nita Farahany led our program. It was with that introduction that I started to think about the fear I have of being in court for a closing argument where someone raises the issue of a deepfake that had not been addressed at all during trial. So I started thinking about what sort of rules we should have in place.

There wasn't a single moment; it evolved through this process. First was the concern about hallucinations. It seemed to me that hallucinations received an outsized, an oversized focus, in that the basic professional responsibility rules cover that problem. Rule 11, and the rules of candor, require that lawyers know and supervise the information they provide to the courts.

So what I did, I initially wrote a simple three-page order. It indicated that people could use generative AI in court; that they had a duty to understand the technology; that they needed to comply with Rule 11 and the other rules of professional responsibility; and that before they challenged something as a deepfake, they needed to give the court notice in advance, pretrial, unless an emergency existed. That simple order received more recognition than I anticipated. I looked around and found that there's a patchwork across the country, and the RAILS program gave me the opportunity to study this in a year-long process.

In thinking about how it should properly be divided, I arrived at what I call the Function Rule. For advocacy purposes—if someone uses AI to write a brief, prepare pleadings, those sorts of things—the regular rules of professional responsibility apply. Where things are different is if someone uses AI to create evidence or affect evidence. Then the court and the opposing side need that information sufficiently in advance to know whether it is reliable and accurate. You can't cite-check evidence the way you can a brief, so advance notice is the fundamental part of the second part of the Function Rule. And finally, the rule reminds people that they're responsible for what they file, argue, or submit to the court.

From there, I prepared a more comprehensive order, published it, and asked for feedback, and I've made some adjustments since, which we can talk about further. So there wasn't really a single moment. The important parts were the program at Duke; preparing an order and submitting it for broader criticism; and finally, the ultimate rule I came up with, which modifies so that it's not just whether AI touches evidence, but whether it materially affects evidence. That, I think, is the distinction. I did not want the rule to be over-inclusive or under-inclusive.

Kim Mueller [00:10:34]: We'll get into the process you used in a little bit, but let's stay on the basic approach. You've mentioned Rule 11 a couple of times. For those listening, you may well

know that I hear Federal Rule of Civil Procedure 11. There's a federal rule, and there's always a state counterpart. Perhaps it's also Rule 11 in North Carolina, Judge McGee?

Judge Marty B. McGee [00:10:53]: It is.

Kim Mueller [00:10:54]: It sets a very high standard for attorneys. Whatever you file with the court, you sign the document and you stand behind it. You're representing to the court, as an officer of the court, that it is accurate and that all the authorities are consistent. So that's a very helpful distinction: advocacy, no disclosure; material effect on evidence, disclosure. Aparna, hearing and knowing what Judge McGee has done here, what possibilities do you see for technologists like yourself when courts do provide clear guidance on AI use? And what clarifications does the profession perhaps still need?

Aparna Komarla [00:11:42]: First and foremost, the biggest value comes from court orders like Marty's, and also SB 574 in California, which require similar things related to human in the loop. They require attorneys to validate that any AI-assisted evidence or material they present in court has been checked for errors and for hallucinations. SB 574 also requires that AI only be used to assist an attorney's legal work, not to replace the intellectual component or decision-making that attorneys do day to day in their legal practice.

One of the biggest gains for us as technologists is that we're not building in the dark. We have very clear do's and don'ts and very clear red lines. It also creates new opportunities to think about how we enable attorneys and other users of legal AI products to do that validation work, how we define what human in the loop really means for them, and how we build systems that are interpretable and explainable and help users comply with a court order like Marty's or with SB 574 in California.

To give you some concrete examples, in my own work at Redo, we're very intentional about which jobs we think AI is good at solving and which tasks AI should not be involved in at all. We know from several research studies that AI does not do a good job at statistical reasoning or computation. It's very good at language tasks: generating narratives, doing synthesis, perhaps explaining a statistical concept. What we're building at Redo is a statistical evidence toolkit that helps attorneys present evidence under California's Racial Justice Act, which requires showing disparities in sentencing outcomes at a systemic level. That evidence can support resentencing, which could mean reconsideration of a past conviction, a sentence reduction, or sometimes elimination [of a sentence].

In doing that, we need to be very careful about which parts of the problem we use AI to solve and which parts we think AI is not good at. When we create an evidence report, we also need to be very clear about which sections have an AI component and which don't, because that ultimately connects to what is required of attorneys and practitioners: disclosing when they're using AI and for what purposes. Overall, having this guidance really helps us think through where those lines are—what we are able to use AI to do, and where we likely should not.

Kim Mueller [00:14:54]: Thank you for the phrase “human in the loop”—HITL, I think you use in your paper. It's important to know that real humans have a role to play in exercising judgment. Just so we're clear: we hear a lot about LLMs today, large language models. You're applying AI to databases where you know what the data is—for example, criminal records and sentencing records for persons serving sentences in California—but you're running an AI tool against them.

How do you think about the LLM that tool was trained on? How selective are you in choosing the right AI?

Aparna Komarla [00:15:42]: That's a great question. We run a number of evaluation studies before we choose which model to use. We have an evaluation rubric, and we recruit human experts—typically statisticians, because the purpose of Redo as a platform is to translate between statisticians and legal practitioners, to make statistical evidence interpretable and actionable. We run a number of these evaluation studies and then use that as a data point to pick which models and which combinations of inputs give us the best output: the best performance and the best accuracy.

But I think the bigger question is what's next for LLMs. Overall, I don't think we'll see huge shifts in the architectures or the kinds of data that large language models are trained on. The onus is really going to be on us as practitioners to determine which tasks we use AI to accomplish, and where we draw red lines and say AI is not good at assisting with these other tasks. There will also be big questions about how we incorporate human expertise. In our work, part of the evaluation is with human statisticians, who measure the outputs of LLMs to make sure they meet a standard a statistician would consider good and worth presenting in court. Those are some of the ways we think through evaluation and model selection.

Kim Mueller [00:17:29]: Thank you for that—emphasizing again the importance of humans in the loop, and careful, informed thinking about what's going on here. Let's move on to our next segment. Judge McGee, did you want to respond to Aparna before we move on?

Judge Marty B. McGee [00:17:46]: If you don't mind.

Kim Mueller [00:17:46]: Not at all.

Judge Marty B. McGee [00:17:47]: This might simplify my proposed rule if I talk about it in context. If a lawyer used the AI system Aparna has created to produce a statistical report to evaluate whether a particular case would be worth pursuing, or to assist in developing arguments before the court, I don't think those uses need to be disclosed. If, however, someone wanted to put an exhibit sticker on a report and introduce it into court, and AI was used in producing it, then I think it's fair that people are aware of that and have an opportunity to decide whether they wish to test it for accuracy, for example.

That does not mean that because AI was used and disclosed, the evidence is admissible or inadmissible. That's a whole other area that would need to be covered. My rules are meant to protect fairness, so there's no unfair surprise. Ultimately, if there were a challenge to it, the court would have to hold a hearing and determine whether the evidence should go before the jury or the fact finder.

Kim Mueller [00:19:17]: Let me ask Aparna: was that all clear to you from Judge McGee's standing order? Would you have understood the clarification he just provided? You have it now.

Aparna Komarla [00:19:25]: Yes, absolutely.

Kim Mueller [00:19:26]: Uh-huh.

Aparna Komarla [00:19:27]: I think there are going to be differing opinions, though, about what kinds of AI-assisted evidence can and should be disclosed, and where to draw the line on

whether a given piece of evidence or report is admissible or not. To be on the safe side, it's always good to make that disclosure and make it known in a report when and where AI is being used. That's something we really encourage on our platform. We're very careful that wherever a section of a report contains analysis that leverages AI, we put the classic AI sparkle next to it, so it's very clear that the section is AI-assisted. Then an attorney can make that judgment call themselves based on the court they're in and what they know about their own judge—to go the extra mile and make the disclosure, or, if they think they don't need to, then they obviously don't have to. But it's helpful for us as technologists to always keep in mind, whether or not the order requires it, and whether in California or North Carolina, to be very transparent about when and how AI is being used.

Testing Ideas in Public

Kim Mueller [00:20:45]: That's music to a former judge's ears: err on the side of disclosure. And Judge McGee makes clear that there will be discovery, so disclose in a way that allows meaningful—hopefully not excessive—discovery to get to the bottom of what might be material, or not.

Before we move on, let me say that I know judges look to judges like Judge McGee to be in the vanguard: to think this through and share something that can either be picked up wholesale or adapted to a judge's own courtroom. So thank you, Judge McGee, for that leadership, and thank you, Aparna, for responding to what the judge has done.

Let's talk about the process. You referenced it, Judge McGee, and we'd like to hear your reflections. You entered your order, but you invited criticism. You went on LinkedIn—a space I'm learning more about. You went to CLEs and Legal Week in New York, and you were on podcasts. That's not how a lot of judges think about getting feedback, if they want feedback. What did testing your ideas in public teach you that you couldn't have learned in chambers? And knowing what you know now, would you do it the same way next time?

Judge Marty B. McGee [00:22:16]: Well, this is how I do lots of things—

Kim Mueller [00:22:19]: Oh.

Judge Marty B. McGee [00:22:19]: —from picking restaurants to crowdsourcing, that sort of thing. I try to maintain my own judgment, but I believe the notion that two heads are better than one. Getting feedback and learning from others is really important, and it helps me get to a better answer. The best answer has to win, regardless of who presents it. That's what I tried to do.

I did a very simple order, and I wanted to seek feedback before I even used it in my own local court, but I thought that process would take too long. So I entered an order, then wrote an article about it and asked for feedback. That was just a wonderful experience, and it got more attention than I expected. On one podcast, I was asked, "Why don't you require people to disclose if they use AI in advocacy?" And I thought, well, I really don't know what I would do with that information. Would I look at that filing more closely? How would that information be helpful? Discussions like that helped me push further and think about what I could do to help move toward more uniformity.

And the RAILS opportunity was just extraordinary. Access to these tools was at least as valuable to me as the experience of drafting the order. In regulating, in trying to come up with the rules, I

thought I should use the tools as lawyers use them, and judges simply don't have access to these tools—or at least very few do, that I'm aware of. So I got them and started out small. I started by using the tools after I'd made a decision, putting the question into the system to see what answer I'd get back. From there, I moved to using them to help with research, and then I used them live in court when I was hearing arguments. I would have done analysis either ahead of time or as the trial went on, so I'd have a response from one of the tools that I could ask the lawyers about. That experience was terrific.

Then I moved to writing on LinkedIn about it and asking for folks' opinions, because I don't have a technology background. I became a judge before e-discovery was a common thing. So I sought out opinions on LinkedIn, and RAILS connected me with others I had conversations with. I built quite a community on LinkedIn. People from all around the world commented on my order, which surprised me—people from Australia, London, and Singapore. It seems people are having this problem all across the globe, and so that feedback was incredibly helpful.

I was trying to move on from that three-page order. I had published a second order that was really too complex, and I wanted to simplify it, and the responses helped. For example, I was trying not to make the rule over-inclusive, and I ended up modifying it in part based on responses and interactions in which I learned that when you take a picture with a simple iPhone, the tools inside the phone can use AI to affect that very picture. So I thought, if my rule were simply "if AI affects evidence," we would have disclosure for every photograph taken with a modern smartphone, and that would not be meaningful information. That helped me modify it in the ultimate model order we published, so that it applies if AI materially affects evidence or is used to create evidence.

I don't think I would have been able to make as much progress as quickly as I did without that process. I also learned an enormous amount about the power of these tools, how they could be helpful to me in chambers, and what some of their limitations are. That was as useful as ultimately coming up with an order, and having a broader perspective.

Kim Mueller [00:27:00]: Thanks for sharing that. Of course, judges can speak outside the courtroom. They need to be aware of the ethical constraints, but thinking about judicial administration—an issue not tied to any one case—is completely permissible and, as you've described, very valuable. There are sometimes court staff, but your description of court processes being deliberative that they take longer than may make sense here is interesting to hear.

Kim Mueller [00:27:37]: We may want to talk more about court infrastructure for thinking about this kind of thing going forward. In the meantime, let's hear from Aparna about the way you evaluate your own tools. You've talked about the importance of transparency. Do you follow anything like the process Judge McGee did? What is your process for getting feedback and refining and improving your products?

Aparna Komarla [00:28:00]: Absolutely. I alluded to some of this earlier, but we have two main ways of doing evaluation. One is intrinsic, where we test the actual technical performance of a system. The other is extrinsic, where we look at how the system interacts with real users, which is more of a human-computer interaction question. Intrinsic evaluation is much easier to do, because we have more control on the data and how we set up our experiments.

We recruit a number of human experts, and we give them a task with an evaluation rubric, and ask them to compare the AI's outputs to the output they would produce themselves, evaluating on the rubric how close or far the AI's output is from their own judgment and expertise. In our case, because we're building statistical evidence toolkits, we recruit statisticians with extensive experience writing, for example, a prima facie showing or assisting with a discovery motion—people with experience in both statistics and computational social science. We ask them to look at the evidence reports we generate and measure several dimensions: how well the model makes sense of, for example, an odds ratio or relative risk; how well it understands statistical significance; and whether it contextualizes the findings adequately given the limitations of the underlying data and then we compare those results to see whether the system is performing well. That gives us data and information to build a feedback loop, and we take it back to the drawing board to think about how to improve the system technically.

Extrinsic evaluation is where things get trickier. We haven't done this formally yet, but we're designing a user study that will combine a survey with semi-structured interviews. We hope to sit down directly with attorneys and legal practitioners and talk with them about their perspective on the platform: how comfortable they feel using it, how comfortable they feel presenting something with an AI component in court, the kind of pushback they anticipate they might get from their judge, and any additional features or mechanisms they'd like in these evidence reports to help them understand how and why AI is being used.

Informally, in conversations with attorneys and practitioners, some of the concerns that come up are more psychological and humanistic, which makes them harder to overcome. We'll get concerns like, “Maybe this is okay to present in court, but I feel nervous doing it, because I don't want pushback from opposing counsel, or I don't want the judge to say the evidence is inadmissible.” Even with a very clear court order like Marty's that says, “This is permissible; you don't need to worry about whether you can use this as evidence, because it falls under such-and-such category,” there's still resistance and hesitance among attorneys about presenting anything with AI in it. That's a much more challenging barrier to overcome.

The way we approach that is: when it does work—when we present an evidence report to an attorney and they get it immediately and say, “Yes, this is safe. We can use this because it meets all of these criteria. It's been validated by a human. There's a human in the loop. There's a very clear handoff or delegation between the humans and the AI. It checks all of the boxes”—we know exactly the kind of language and framing of the product that can help alleviate skeptics' concerns. That's how we're thinking about evaluation overall. Our hope is that all of these data points—from formal experiments, user studies, and informal conversations—will help us design these systems better.

Kim Mueller [00:32:38]: Maybe I missed it: have you gotten feedback from any judges so far?

Aparna Komarla [00:32:43]: Not from judges.

Kim Mueller [00:32:45]: Okay. So it's—

Aparna Komarla [00:32:45]: Just the attorneys.

Kim Mueller [00:32:46]: Okay. So that's to be determined, and it will be interesting to see how it plays out. In a common law system, things will be decided case by case and sorted out over time. But the care you're taking is very impressive. We keep talking about Judge McGee's order,

and I see his hand. Just so it's clear, it's a standing order: a judge can issue an order that applies to every case, with the same rules, and that's what Judge McGee has done here. Judge McGee, your response to what you've heard, or anything else you'd like to say on this topic?

Judge Marty B. McGee [00:33:26]: I was interested, from Aparna's perspective: does she intend the report itself to be presented along with an expert, as part of the expert's testimony explaining the analysis? Or do you intend it to be admitted as a standalone exhibit without an expert explaining it?

Aparna Komarla [00:33:52]: That's a great question. We're open to any and all use cases. When we produce these reports, we tell attorneys they can use it simply as an exploratory analysis tool. Maybe they're thinking about statistical evidence and want to experiment with Redo and a couple of other similar products in the public domain, such as the Paper Prisons project at UC Berkeley, to help them think through evidence they could create. So that is one way they might use it, just for exploratory analysis. Ultimately, they may produce a report that takes inspiration from the outputs of the various tools they've consulted, but that report is materially different from the Redo report. That's one way it could be used.

Another way a report could be used is for a prima facie showing or in a discovery motion; for instance, an attorney could present the report as a standalone exhibit. A third way: if I were retained as an expert by a public defender's office, I might take the Redo report and present it alongside my testimony. Those are different ways the report—the statistical evidence plus the AI interpretation—can be presented, and all of them will require different disclosure and different standards for admissibility, and will have to be treated differently based on how it's being utilized.

Kim Mueller [00:35:34]: Any further thoughts on that, Judge McGee?

Judge Marty B. McGee [00:35:36]: I think that's interesting. Expert reports are one place where AI could come into evidence, and my proposed rule specifically discusses when counsel has a duty to disclose an expert's use of AI. That does not mean the report is inadmissible—just that AI was part of the analysis. A statistical report presented through an expert would likely be closer to the way things normally go. Going a step beyond that, to say that an AI-created report would itself be admissible into evidence, is another area and a step further. It will be interesting to see how the rules catch up to that and whether it would be permissible now or in the future. The admissibility issue is outside my particular order, but I think it's an interesting subject.

Educating the Profession

Kim Mueller [00:36:51]: People you mentioned earlier, Judge McGee—my predecessor, former District Judge Paul Grimm, and Maura Grossman—are thinking deeply about evidence rules and whether we need more there or whether the existing rules can work. I see the questions coming in. We'll cover one more section, have a wrap-up question, and then get to your questions. Kelli, at some point perhaps you can let us know in the chat: I think the materials we're referencing—Aparna's paper and Judge McGee's order—can be found on the RAILS website. I was able to access them, and we want to make sure you can all find them. Kelli is confirming that it's all on the website. [rails.legal/fellows]

Let's move to educating the profession, and let's start with you, Aparna. Given your background in data science and the way you've laid it out for us today, what do you most want lawyers and

judges to understand at this point about the tools that are currently available and that you're working to develop and refine?

Aparna Komarla [00:38:06]: A couple of things come to mind. As technologists, we are sometimes not very careful with details or traceability. We skip over a lot of steps when we explain how a system works. I've gotten a lot of really good advice from my own mentors on how to describe the product and how AI is being used, and one of the most useful pieces of advice is that no detail is too small. Be very thorough about even the most minute step: button A is clicked, followed by button B, followed by button C. That level of detail in describing methodology is very helpful for building trust, increasing transparency, and really answering the "how" question.

I would also say that AI has a lot of value, and there's so much excitement about this technology and what it can do. The more orders we have like Marty's, and the more laws we have like SB 574 in California that dictate what can and can't be done, the better we are placed in mitigating algorithmic harms. Algorithmic harms existed before LLMs, too. For those who are familiar, there was a lot of outcry about COMPAS and pretrial risk assessments in the criminal legal system because, according to some studies, they showed bias against minority and disadvantaged groups. There was a lot of back and forth in the computer science community for quite some time, because tools like COMPAS didn't all use race as a variable. We thought you were safe if you didn't use race as a predictor or variable in your model. But many data points in the datasets acted as proxies for race—a ZIP code, for example. That was one way racial harms were being propagated through a system we thought was unbiased.

One of the really important takeaways for me is that saying no to AI altogether is not going to solve the problem, because people are using it, and we want to make sure they're using it well. An experience I had recently really crystallized this for me. Under the RJA, the first step for attorneys is grouping together what are called similarly situated cases: the law requires showing statistical disparities in sentencing outcomes for folks who are similarly situated. An attorney I was talking with asked, "If I have a list of 1,000 people, could I plug that into a chatbot like Claude or GPT and ask it to give me the top 10 cases that are most similar to each other?"

Aparna Komarla [00:41:21]: I was very concerned hearing that, because there are two big problems. First, these chatbots are black-box models. You don't actually know how they work or what reasoning or logic they use to determine similarity or dissimilarity. Second, you could do validation and human-in-the-loop review after the model returns those top 10 cases, which might help you catch false positives. But what about all of the false negatives? What about the 990 cases the model said were dissimilar? How do you validate those? If we don't have good guidance and laws dictating how AI can be used, we may end up in a world with these bad use cases, because people aren't being educated about the right way to use this technology. For me, one of the big next steps is figuring out how to take all of this guidance and make it truly actionable in the systems we build.

Kim Mueller [00:42:30]: It's so good to hear about the value of orders like Judge McGee's—performance standards that really guide the thinking of people like you who are thinking so carefully about this. Judge McGee, stepping back, from the bench: what more do you want tool builders like Aparna to know about AI and litigation, given where you are in your thinking about all of this?

Judge Marty B. McGee [00:42:55]: I think it would be very helpful if tool builders gave judges and court personnel more access to the tools. It's very difficult to understand what they can do and what their limitations are without using them. I would also say there are opportunities beyond legal research. A public-facing AI tool used by a court could be developed to greatly increase access to justice; I've been working on a side project on that, and I think it could be incredibly helpful. There are lots of things that can be automated where human judgment doesn't need to be individualized, while still protecting people's right to come to court.

I'd like to think about tools being built not just for courts, but with courts, with courts participating. Toolmakers and people in the legal profession working together would be helpful. The people on the ground making these decisions, who understand the process, could work with technologists to create better tools to help courts, lawyers, and the public, and to increase access to the courts. I think there's a lot of promise there.

Advice for Judges

Kim Mueller [00:44:25]: Thank you, Judge McGee, for pointing that out. The Bolch Judicial Institute is committed to facilitating exactly the kind of convening you've just described to improve access to justice, including with international justice actors. That's one area where we hope to keep working with you and the folks at RAILS, because we agree that we should be able to improve access to justice if we think it through and know what the tools are.

Before we go to questions, which Kelli will facilitate, let's think briefly about lessons for any judges who may be listening—I know we have some—or people who work within court systems. We'll start again with you, Aparna. What would you say to a judge whose instinct is to flatly prohibit AI tools until all the risks are better understood, if not eliminated? I think we may have a sense of what you'd say, but what would you say to this audience today?

Aparna Komarla [00:45:33]: As I described earlier, I think AI is here to stay, and every field is going through its own transformation. In the data science world I come from, for example, we're seeing interesting cleavages in the kind of work we do. There are tasks that will almost certainly be human-AI collaborative, tasks that will be fully automated and done by AI, and tasks that will be retained by humans. Every field is seeing this kind of transformation, and naturally the legal field will go through this journey as well.

Banning AI outright or putting a moratorium on it will only perpetuate the kinds of bad use cases where folks have incomplete knowledge of how a system works, with a mental model that might not reflect how the system actually operates under the hood. With good education about these tools, and with court orders like Marty's and laws like SB 574, we can really steer toward good use cases.

I also really like what Marty said earlier about having as much openness as possible in the platforms we build. That's hard to do, because many models today are closed. GPT is a closed model, and Claude is a closed model, meaning the underlying weights are not open to the public, and we don't see all of the training data that goes into these systems. There are small language models and open-weight models, but those are not as well suited to some of these legal tasks. Overall, there are things we can do as technologists to be more transparent and open about how our systems work, and to give judges more access so they can experiment with the tools and get comfortable with them. In terms of next steps, being more open, and using these laws and

guidance to steer both technology development and the education of practitioners on the ground, is where I think we should be headed.

Kim Mueller [00:48:03]: Judges will also start learning when there are evidentiary disputes and pretrial hearings over experts: can an expert explain what happened inside a black box? That's the other way judges will learn. Judge McGee—

Judge Marty B. McGee [00:48:17]: Yes.

Kim Mueller [00:48:17]: —just briefly, your guidance to judges generally, in addition to what you've already shared, but also for a judge talking to clerks and staff in chambers. Could you reflect on how should judges approach this? Are outright bans okay? Some flexibility with disclosure? What should judges say to the people they work with and trust day to day?

Judge Marty B. McGee [00:48:42]: First, I would say I understand courts' reluctance to use artificial intelligence. Courts are responsible for important decisions. I don't think, however, that this technology is going away. I don't think anyone is advocating turning human judgment over to a computer. I think we need clear rules: public-facing rules for when lawyers can use these tools and when pro se folks [self-represented litigants] can use them, and also in-house guidance about when it is and isn't appropriate to use them. From what I've heard, very few chambers have specific rules about when AI can be used. I know many law clerks, like judges, don't have access to these tools, so they're very limited. We need to develop that. There need to be clear rules about when AI can and cannot be used.

I would also say we ought to have some grace, because this is new technology that changes every day and that we're all trying to learn. I don't think anyone is including 17 hallucinated cases in their brief because they think they'll slip it past someone and it will help them win the case. I think that's a simple mistake. That mistake probably shouldn't be made now, with as much information as is out there, but there will be smaller mistakes going forward, and having some understanding of that is also important. But clear rules, I think, are important.

Audience Questions

Kim Mueller [00:50:15]: Yes—a fundamental principle for any rule-of-law system: being clear and applying the rules equally. Let's move to some questions. Kelli, would you like to pose the questions, or should I read them?

Kelli Raker [00:50:35]: I can read the questions; that's not a problem.

Kim Mueller [00:50:37]: That would be great. I know you've been monitoring them closely. I see the court reporter question, which is an important one, so let's make sure we have time for it, but let's start with some of the others first.

Kelli Raker [00:50:51]: Sure. One question is whether Judge McGee's published model order has any binding effect on other courts—whom would it potentially affect if it were adopted?

Judge Marty B. McGee [00:51:03]: No. It is simply a suggested order, a starting point for courts to consider. There's a patchwork across the country, and this is my effort to distill some principles that could be universally applicable, or at least a starting point. So no, it's not binding. I have a local order for people who appear in Cabarrus County Superior Court that relates to our courthouse. But my model order is, I hope, a starting point that will continue the conversation

about how we can move toward more uniformity, which I think will benefit lawyers, judges, toolmakers, pro se [self-represented] litigants—everyone.

Kim Mueller [00:51:47]: It's not binding until a judge issues it. I don't know whether it would ever be incorporated into formal rules after a lengthy process. But if I were still a judge, I could look at Judge McGee's order and make it the standing order in my courtroom, and then it would be binding on all of the cases filed before me. Unless or until a judge adopts it that way, it has no binding effect. But it's out there, and Judge McGee has offered it to the world, so please feel free to consider adopting or adapting it in your own court if you're a judge. Kelli, next question.

Kelli Raker [00:52:29]: Along those lines: who should create the rules surrounding AI? The state legislature, chief judges, the federal government? Who else should be involved?

Kim Mueller [00:52:40]: Good question. Aparna, do you have thoughts on that? These are rules to apply in court cases; let's narrow it that way.

Aparna Komarla [00:52:52]: There's a lot of development around the Federal Rules of Evidence, and I saw another question in the chat about FRE 707. For us as technologists, there's still some ambiguity about exactly what we're permitted to do with respect to the three ways an AI-assisted report can be presented in court that we talked about: presenting the report as is without an expert witness, presenting it with an expert witness, or an attorney using it for exploratory analysis.

The way we've approached this is by creating space for different kinds of use cases in a very practical sense. For example, the reports our system generates are actually Word documents. That was a very intentional design decision. It sounds simple, but it's really important: if there's an AI component an attorney feels less confident about because of one of these rules of evidence, they can take it out of the report altogether, clean it up, and present only the non-AI components of the statistical analysis. Because some of these laws are still being decided, that's how we've tried to make room for different ways the systems can be used. I don't have more to add on that.

Kim Mueller [00:54:32]: Judge McGee, your thoughts?

Judge Marty B. McGee [00:54:34]: I believe process rules are local. Federal judges and state judges enter administrative or standing orders concerning the process before them. Our state legislature creates the rules of evidence, the civil procedure rules, and those sorts of things, and committees obviously participate in that. Congress and the state legislatures also have a role. And of course, the appellate and supreme courts of each state, and the federal courts, have a role in promulgating rules of professional responsibility. So it's multi-level. My focus on process resulted in a local order for local decisions.

Kim Mueller [00:55:28]: Courts have processes, and I think over time you will see the rules for each state and the federal rules adapt, but it's a very deliberative process. In terms of the rules of evidence, there's a proposal pending to address AI in the evidence context, though I think it's currently on hold. In the meantime, the vanguard—the Judge McGees of the world—will do what they can to provide clarity, and things will evolve.

Kim Mueller [00:56:03]: Courts' management of themselves involves thinking carefully about the rules. For the federal rules, the courts submit them to Congress, and very often they become

final without any action by Congress. All right, Kelli, there may be time for one or two more questions. Is there another before the court reporter question, or should we—

Kelli Raker [00:56:23]: We can jump to the court reporter question. That's a good one to go to next. It reads: "As a traditional stenographic court reporter, I and my colleagues are very concerned about new so-called court reporting agencies that simply record proceedings and then use AI to create transcripts. We just don't believe AI technology is accurate enough. I wonder what you all think and feel about AI's use in creating the record, and what you would suggest in the way of regulation of that practice."

Kim Mueller [00:56:52]: I have some thoughts, but Judge McGee may be the one to answer this. Aparna, I don't know whether you have anything to say on this—probably not.

Aparna Komarla [00:57:00]: I'm curious to hear what Marty says.

Kim Mueller [00:57:02]: Okay. Judge McGee, and then I have thoughts. That was an issue in my court before I retired.

Judge Marty B. McGee [00:57:10]: This is obviously a developing area, and I have touched on it in my model order. First, I have concerns about people recording in court and whether it's appropriate for anyone other than the court reporter to do that. That's a separate but related issue.

I'll give an example. I was presiding over a homicide trial. It wasn't courtroom testimony, but the prosecutor had a recorded jail call, which was sent to the Administrative Office of the Courts, and, as I understand it, a purely AI-generated transcript was created without involving a court reporter. There was an objection in court. The prosecutor wanted to play the audio and give the transcript to the jury to help them follow along. No one had raised that in advance, so I had to send the jury out and figure it out on the spot. We went through it, and there were errors in it. I asked the lawyers, "Are they material errors?" and we went back and forth. Ultimately, I decided the errors were not material, and I allowed the transcript to be used for that purpose.

Whether we move to having AI record all sorts of hearings is a developing issue. For trials in North Carolina, the preference is to use a live court reporter. In some places, because we lack court reporters, recordings are used. If a transcript is purely AI-generated and has not been certified by a court reporter, I think that should be brought to the court's attention and disclosed. Ultimately, the court will have to make a determination whether it is admissible.

Kim Mueller [00:59:19]: I would just add that this is such an important issue to think about. There has to be one court record. For any court proceeding, there cannot be competing records of what happened in court. Judges often make clear that no one else gets to record; there is only the court reporter's transcript. Whether AI becomes more common is tied to the availability of live court reporters. In my court, there often were not sufficient humans in the loop—human court reporters—and so judges would turn to AI tools. In my mind, there's no replacement for the human court reporter, who can not just take down the words but also create a transcript that lets you read it and recreate, three-dimensionally, what happened in the courtroom. Some may say AI is heading in that direction, but—

Kim Mueller [01:00:22]: —court reporters can say, "During this pause, the witness is sobbing on the stand," or "The attorney is waving a gun in the face of the witness." So I'm a strong believer in the value of live court reporters. I think court reporters would do well to ensure there

are people coming through the pipeline and being trained to serve in that capacity, and also to figure out how they can use AI to expedite the preparation of transcripts—but certify them only once a human has checked them, compared what the transcript says to what actually happened in court, and signed off. Those records are so important, so thank you to the court reporter who's been watching, for what you do.

We are at our time, so we'll wrap up with great thanks to Judge McGee, Ms. Komarla, the RAILS program, and all of you for listening and posing your questions. I know I have learned tremendously from this, and I look forward to reading the transcript, which we'll make available, and to continuing to think about the important issues we've raised today. Thank you all. Thank you to Kelli and the RAILS program, our partner in offering this. Have a good rest of the day. Take care.