The GPTJudge: Justice in a Generative AI World

Maura R. Grossman, et. al.

23 Duke L. & Tech. R. 1 (2023).

THE GPTJUDGE: JUSTICE IN A GENERATIVE AI WORLD

MAURA R. GROSSMAN, HON. PAUL W. GRIMM (RET.), DANIEL G. BROWN, AND MOLLY (YIMING) XU[†]

ABSTRACT

Generative AI ("GenAI") systems such as ChatGPT recently have developed to the point where they can produce computer-generated text and images that are difficult to differentiate from human-generated text and images. Similarly, evidentiary materials such as documents, videos, and audio recordings that are AI-generated are becoming increasingly difficult to differentiate from those that are not AI-generated. These technological advancements present significant challenges to parties, their counsel, and the courts in determining whether evidence is authentic or fake. Moreover, the explosive proliferation and use of GenAI applications raises concerns about whether litigation costs will dramatically increase as parties are forced to hire forensic experts to address AI-generated evidence, the ability of juries to discern authentic from fake evidence, and whether GenAI will overwhelm the courts with AI-generated lawsuits, whether vexatious or otherwise. GenAI systems have the potential to challenge existing substantive intellectual property ("IP") law by producing content that is machine, not human, generated, but that also relies on

---

[†] Maura R. Grossman, J.D., Ph.D. and Daniel G. Brown, Ph.D., are professors, and Molly (Yiming) Xu is an undergraduate student (as well as Drs. Grossman and Brown's research assistant) in the David R. Cheriton School of Computer Science at the University of Waterloo. Dr. Grossman is also an adjunct professor at Osgoode Hall Law School of York University and an affiliate faculty member of the Vector Institute of Artificial Intelligence. Hon. Paul W. Grimm (ret.) is the Director of the Bolch Judicial Institute and the David F. Levi Professor of the Practice of Law at Duke Law School. Previously, he served as a District Judge (and before that as Magistrate Judge) in the United States District Court for the District of Maryland. Drs. Grossman and Brown's work is funded, in part, by the National Science and Engineering Council of Canada ("NSERC"). The authors wish to thank Katherine Gotovsky, Amy Sellars, Gordon V. Cormack, and Hon. John M. Facciola (ret.) for their thoughtful comments on a draft of this paper; their comments helped us to clarify and strengthen some of our arguments. The views expressed in this article are the authors' own, and do not necessarily reflect the opinions of the institutions with which they are affiliated.

human-generated content in potentially infringing ways. Finally, GenAI threatens to alter the way in which lawyers litigate and judges decide cases.

This article discusses these issues, and offers a comprehensive, yet understandable, explanation of what GenAI is and how it functions. It explores evidentiary issues that must be addressed by the bench and bar to determine whether actual or asserted (i.e., deepfake) GenAI output should be admitted as evidence in civil and criminal trials. Importantly, it offers practical, step-by-step recommendations for courts and attorneys to follow in meeting the evidentiary challenges posed by GenAI. Finally, it highlights additional impacts that GenAI evidence may have on the development of substantive IP law, and its potential impact on what the future may hold for litigating cases in a GenAI world.

INTRODUCTION

In the past few months, generative artificial intelligence ("GenAI")—deep learning models that can produce text, images, and other content based on their training data—has come to the forefront of the news media and captivated the public's attention. Students are using OpenAI's ChatGPT to do their schoolwork, to the alarm of teachers and school boards.[1] An administrator at Vanderbilt University used ChatGPT to write a message to the university community in response to tragic shootings at Michigan State, which sparked outrage.[2] Websites are using images generated by

---

[1] Rob Waugh, *'Half of School and College Students Are Already Using ChatGPT to Cheat': Experts Warn AI Tech Should Strike Fear in all Academics,* DAILY MAIL (Mar. 26, 2023, 13:41 EDT), https://www.dailymail.co.uk/sciencetech/article-11899475/Half-students-using-ChatGPT-cheat-rise-90.html; Arianna Johnson, *ChatGPT in Schools: Here's Where It's Banned—And How It Could Potentially Help Students*, FORBES (Jan. 31, 2023, 11:32 AM EST), https:forbes.com/sites/ariannajohnson/2023/01/18/chatgpt-in-schools-heres-where-its-banned-and-how-it-could-potentially-help-students/?sh=2b5bb4f76e2c.

[2] Sam Levine, *Vanderbilt Apologizes for Using ChatGPT in Email on Michigan Shooting*, THE GUARDIAN (Feb. 22, 2023, 08:58 EST), https://www.theguardian.com/us-news/2023/feb/22/vanderbilt-chatgpt-ai-michigan-shooting-email.

Midjourney[3] and Stable Diffusion,[4] and cover artists and other illustrators are suddenly fearing for their livelihoods.[5] Clarkesworld, a major science fiction magazine, had to close its doors to new submissions, after an influx of AI-generated stories prevented it from performing its normal review process for new manuscripts.[6] Increasingly lifelike pornographic videos and still images are created using AI systems that incorporate the faces and bodies of celebrities and other pop culture figures into the media they are generating.[7]

These systems did not come out of nowhere. Systems that simulate creativity or that generate text have been a thriving branch of computer science research for decades. But in the past few years, this technology has become so powerful that it is now challenging to tell computer-generated images from those produced by human illustrators or photographers,[8] or to separate computer-generated text from that written by human authors.[9] Similarly, evidentiary materials—including documents, videos, audio recordings, and more—that are AI-generated are becoming increasingly difficult to distinguish from those that are non-AI generated.

---

[3] MIDJOURNEY, https://www.midjourney.com/home/?callbackUrl=%2Fapp%2F (last visited Oct. 16, 2023).
[4] STABLE DIFFUSION, https://stablediffusionweb.com/ (last visited Oct. 16, 2023).
[5] Rob Salkowitz, *AI Is Coming For Commercial Art Jobs. Can It Be Stopped?*, FORBES (Sept. 16, 2022, 02:10 PM EST), https://www.forbes.com/sites/robsalkowitz/2022/09/16/ai-is-coming-for-commercial-art-jobs-can-it-be-stopped/?sh=3bc8d48b54b0.
[6] Alex Hern, *Sci-fi Publisher Clarkesworld Halts Pitches Amid Deluge of AI-Generated Stories*, THE GUARDIAN (Feb. 21, 2023, 14:27 EST), https://www.theguardian.com/technology/2023/feb/21/sci-fi-publisher-clarkesworld-halts-pitches-amid-deluge-of-ai-generated-stories.
[7] Moira Donegan, *Demand for Deepfake Pornography Is Exploding. We Aren't Ready for this Assault on Consent*, THE GUARDIAN (Mar. 13, 2023, 06:16 EDT), https://www.theguardian.com/commentisfree/2023/mar/13/deepfake-pornography-explosion.
[8] *See, e.g.*, Simon Ellery, *Fake Photos of Pope Francis in a Puffer Jacket Go Viral, Highlighting the Power and Peril of AI*, CBS NEWS (Mar. 28, 2023, 11:39 AM EDT), https://www.cbsnews.com/news/pope-francis-puffer-jacket-fake-photos-deepfake-power-peril-of-ai/ (describing the dangers of AI).
[9] *See* Jan Hendrik Kirchner et al., *New AI Classifier for Indicating AI-Written Text*, OPENAI (Jan. 31, 2023), https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text (explaining the limitations of OpenAI's new classifier).

While it may seem like GenAI's appearance in the courtroom is a distant prospect, judges should not be lulled into false complacency. These cases will be coming their way much sooner than they think, and the courts need to be ready for them. By way of example, imagine the following scenarios.

I. COMING SOON TO A COURT NEAR YOU

Several days before beginning her final undergraduate semester, Keisha, a pre-law student at Georgetown University, received a devastating email from the Dean's Office accusing her of cheating on her political science honors thesis during the preceding semester. The work in question was an essay that Keisha had submitted concerning U.S. federal government policy regarding biometric data collection, which she had written with the help of ChatGPT, a GenAI tool that responds to dialogue-styled prompts with narrative text.[10] Keisha responded to the email arguing that under the University's academic guidelines, writing with the unauthorized help of another *person* would be considered cheating, but there were no rules prohibiting other forms of assistance, such as artificial intelligence, and that she had both personally prepared the prompts provided to ChatGPT and reviewed the final work product that was submitted. The University also disciplined Keisha on another ground: She had fabricated material and attributed it to a real source. Although Keisha had proofread and edited the essay produced by ChatGPT, she did not cross-check all of the references because ChatGPT cited the sources with such authority; it never occurred to her that they might be faulty AI "hallucinations."[11] After having been rejected on all her law school applications—ostensibly as a result of the failing grade on her thesis and the violation of Georgetown's academic integrity rules—Keisha sued the university. In her complaint, she alleges that her friend, who is not a native

---

[10] *Cf.* Pranshu Verma, *A Professor Accused his Class of Using ChatGPT, Putting Diplomas in Jeopardy*, THE WASHINGTON POST (May 18, 2023, 2:19 PM EDT), https://www.washingtonpost.com/technology/2023/05/18/texas-professor-threatened-fail-class-chatgpt-cheating/ (explaining that AI-generated writings can be almost impossible to detect in classrooms).

[11] *See* Ziwei Ji et al., *Survey of Hallucination in Natural Language Generation*, 55 ACM COMPUTING SURVEY 1 (2022) (discussing the tendency of Natural Language Generation models to provide false information).

English speaker, has routinely used tools like spellcheck and Grammarly,[12] and has never been disciplined for receiving unauthorized assistance. One of Keisha's claims is that the distinction between what she did and what the other student did is unfair and discriminatory. Keisha's case has been assigned to you.

Sam is a freelance artist who works with many different forms of digital media. Recently, he noticed that several of his friends had changed their online profile photos to drawings of themselves and he decided to do the same. While scrolling through TikTok, he noticed a familiar drawing in a video about an app that could transform photographic selfies into drawings. If it were not for the remnants of a blurred logo at the top right corner, Sam might not have been able to confirm that this AI-generated drawing was based on a sketch he had posted online a few years earlier. After discussing his experience with other artists in his local community, Sam realized that this trend could threaten the livelihoods of many artists other than just himself. The app in question integrated DALL-E 2,[13] which can create unique images using training datasets taken—without consent—from artists' work found on the Internet. Using this as a starting point, Sam and a coalition of artists sued several GenAI companies with similar AI models, alleging copyright infringement. The suit includes as defendants not only the companies that built the AI models, but also the companies that collected the data and trained the GenAI algorithms, the company that developed the app he visited, and the individual who made the TikTok video that contained his artwork. This is a case of first impression in your district because, to date, there has been no precedent on whether training on Sam's and his colleagues' data reflects "fair use,"[14] nor any case that addresses who might be liable

---

[12] GRAMMARLY, https://www.grammarly.com/ (last visited Sep. 19, 2023).

[13] *DALL-E 2*, OPENAI, https://openai.com/product/dall-e-2 (last visited Sep. 19, 2023).

[14] Under U.S. copyright law, "fair use" permits the unlicensed use of copyright-protected work under certain circumstances, such as in some non-commercial or educational contexts, including news reporting, teaching, and research. The issue of fair use of prior photographs in subsequent graphic art was addressed by the Supreme Court on May 18, 2023, in *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith et al.*, 143 S.Ct. 1258 (2023). In its opinion, the Court ruled 7-2 that Warhol's reliance on one of Goldsmith's photographs of Prince as an "artistic reference" point in his series of 16 silk-screen images of the musician (known as "the Prince Series") infringed on Goldsmith's copyright and was not

under these facts.[15] The case has been assigned to you.

The elderly have long been easy targets of telephone scams and phishing emails, but GenAI adds a whole new dimension to this problem. Barb, 81, and Henry, 84, are residents of a nursing home in Florida. They recently received an urgent voicemail message appearing to be left by their grandson, Adam, a graduate student at the University of Minnesota. In the message, Adam explained that he was returning home from a party the night before when he was arrested for driving while intoxicated. He stated that he was being held in jail and needed money for bail and to hire an attorney. He pleaded with his grandparents to wire him $12,000. After they received the message from Adam, Barb and Henry listened to it again with a nursing home administrator, who helped them call their bank to arrange for the transfer of $12,000. Adam has a YouTube channel where he posts instructional videos on craft beermaking. It turns out that a scammer entered Adam's voice from some of his YouTube videos into Murf.AI,[16] an AI voice-cloning tool, and was able to convincingly synthesize his voice to defraud his grandparents.[17]

---

fair use because Warhol did not sufficiently transform Goldsmith's original photograph in his derivative work. *Id.* at 1287. The dissent wrote that the majority's decision "will stifle creativity of every sort. It will impede new art and and music and literature. It will thwart the expression of new ideas and the attainment of new knowledge. It will make our world poorer." *Id.* at 1312. Many commentators believe that this outcome could have a profound impact on copyright law; in particular, it could affect the extent to which GenAI systems that rely on copyrighted images infringe on copyright holders' rights. *See*, *e.g.*, Paul Szynol, *The Andy Warhol Case That Could Wreck American Art*, THE ATLANTIC (Oct. 1, 2022), https://www.theatlantic.com/ideas/archive/2022/10/warhol-copyright-fair-use-supreme-court-prince/671599/ (arguing that the outcome of the case could have severe consequences on creativity in the United States).

[15] *See*, *e.g.*, Complaint, Getty Images (US), Inc. v. Stability AI, Inc., No. 1:23-cv-00135-UNA (D. Del. Feb. 3, 2023); Complaint, Anderson, et al. v. Stability AI Ltd., et al., No. 3:23-cv-00201 (N.D. Cal. Jan. 13, 2023).

[16] *Voice Cloning Product Page*, MURF.AI, https://murf.ai/voice-cloning (last visited Oct. 22, 2023).

[17] *See*, *e.g.*, Pranshu Verma, *They Thought Loved Ones Were Calling for Help. It Was an AI Scam.*, THE WASHINGTON POST (Mar. 5, 2023, 6:00 AM EST), https://www.washingtonpost.com/technology/2023/03/05/ai-voice-scam/ (documenting how scammers are using artificial intelligence to sound like family members); *see also* Gene Marks, *It Sounds like Science Fiction but It's not: AI*

Finally, Maria is an undocumented immigrant living in the Bronx, New York. Her baby has been colicky for a few days in a row and appears to be growing increasingly distressed. Maria does not want to go to the local hospital emergency room because of her immigration status and lack of insurance. Instead, she logs on to a search engine that has been augmented with a chatbot feature that uses a large language model ("LLM") and describes the baby's symptoms. The algorithm does not show Maria any pre-existing webpages, rather, it automatically generates an English narrative response to her specific query. In her case, the response suggests giving the baby an aspirin and indicates that the baby should be fine in the morning. However, the baby becomes severely ill the next morning and develops a fever of 104 degrees. Maria rushes to the closest emergency room with her baby. The baby eventually recovers, but Maria is told that the baby will have a long-term cognitive disability because of the delay in receiving appropriate medical treatment. Maria sues the creator of the search-engine algorithm, arguing that it bears responsibility for the advice she received. If the company had merely linked to existing web pages, arguably it would have avoided any liability under Section 230 of the Communications Decency Act of 1996.[18] But, because the search engine provided Maria with a single narrative response (rather than providing a series of links), Maria's counsel argues that it is responsible for damages. The search-engine company argues that because the chatbot feature contains a warning and disclaimer concerning its accuracy, Maria should have realized that the response was not authoritative and therefore, she could not reasonably rely on it. Moreover, because the chatbot was trained on a large dataset of existing Internet information that the search-engine

---

can Financially Destroy your Business, THE GUARDIAN (Apr. 9, 2023, 12:00 EDT), https://www.theguardian.com/business/2023/apr/09/it-sounds-like-science-fiction-but-its-not-ai-can-financially-destroy-your-business (explaining how scammers used AI to steal $11 million by fabricating the voice of loved ones); Joseph Cox, *How I Broke Into a Bank Account with an AI-Generated Voice*, VICE (Feb. 23, 2023, 11:44 AM EST), https://www.vice.com/en/article/dy7axa/how-i-broke-into-a-bank-account-with-an-ai-generated-voice (explaining how the author used an AI-generated voice to break into a bank account).

[18] 47 U.S.C. § 230(c)(i); *see also Section 230*, THE ELECTRONIC FRONTIER FOUNDATION, https://www.eff.org/issues/cda230 (explaining the function and purpose of Section 230).

company did not create, it claims that it is not responsible for damages.[19] The case has been assigned to you. How do you determine who is responsible for what?

These examples are not far-fetched and raise novel and complex issues with which the courts will soon have to grapple.

II. WHAT IS THIS STUFF AND WHERE DID IT COME FROM?

Algorithms for simulating creativity have long been a natural interest of computer science researchers. The mathematical properties of music and language have been a focus of this area; researchers have attempted to reproduce the vocabulary and style of existing composers and authors, and even to use computers to derive entirely new styles of artistic work.[20] Over time, these methods have moved on to other media: video, visual art, animation, and more. They have also intersected with the same technology used to make

---

[19] There has already been at least one lawsuit brought in response to defamatory statements made by Chat-GPT. *See*, *e.g.*, Cassandre Coyer, *ChatGPT Made Up Sexual Harassment, Bribery Charges About Users. Can It Be Sued?*, ALM (May 9, 2023, 6:57 PM EDT), https://www.law.com/legaltechnews/2023/05/09/chatgpt-made-up-sexual-harassment-bribery-charges-about-users-can-it-be-sued/; Rebecca Cahill, *OpenAI Defamation Lawsuit: The first of its kind*, SYRACUSE L. REV. (June 22, 2023), https://lawreview.syr.edu/openai-defamation-lawsuit-the-first-of-its-kind/. Many commentators—including the two congressional leaders who co-authored the law—do not believe that Section 230 will serve as a successful defense for AI-powered chatbots that defame because they do not merely supply third-party content, but rather, they generate new information. *See* Cassandre Coyer, *ChatGPT Faces Defamation Claims. Will Section 230 Protect AI Chatbots?,* ALM (May 22, 2023), https://t.ly/fEn-N; Yisroel Mirsky & Wenke Lee, *The Creation and Detection of Deepfakes: A Survey*, ARXIV:2004.111138 [cs.CV] (Sept. 13, 2020), https://doi.org/10.48550/arXiv.2004.11138 (describing a number of these techniques).

[20] *See*, *e.g.*, Simon Colton & Geraint A. Wiggins, *Computational Creativity: The Final Frontier*, 242 FRONTIERS A.I. 21 (2012) (providing an overview of the field of computational creativity); Kemal Ebcioğlu, *An Expert System for Harmonizing Chorales in the Style of J. S. Bach*, 8 J. LOGIC PROGRAMMING 145, 145 (1990) (describing the functions of "a knowledge-based expert system for harmonization and hierarchical voice leading analysis of chorales in the style of J. S. Bach"); PAMELA MCCORDUCK, AARON'S CODE: META-ART, ARTIFICIAL INTELLIGENCE, AND THE WORK OF HAROLD COHEN (W.H. Freeman & Co. 1990); MARGARET A. BODEN, ARTIFICIAL INTELLIGENCE AND NATURAL MAN, 298–344 (Basic Books, Inc., 2nd ed. 1987).

deepfakes.[21] Not only can contemporary algorithms make movie clips in the style of a famous director, they can also incorporate the realistic likenesses of particular Hollywood stars into that video, having these simulated actors say things the real actors never said.

Better training techniques and more sophisticated content-generation methods have revolutionized these algorithms over the past few years to accurately represent the underlying properties of human-generated base materials. These improved algorithms are often referred to a as "deep learning" methods.[22] Other major developments include the massive decline in costs for collecting and storing training data, and improved technology for building huge training data sets.[23]

Generative AI is a specific subset of AI used to create new content, or to replicate an artistic style, by training on existing data taken from massive data sources—primarily the Internet—in response to a user's prompt.[24] The prompt, and the new content or

---

[21] *See, e.g.*, Sebastian Berns et al., *Automating Generative Deep Learning for Artistic Purposes: Challenges and Opportunities*, PROC. OF 12TH INT'L CONF. ON COMPUTATIONAL CREATIVITY (2021), https://computationalcreativity.net/iccc21/wp-content/uploads/2021/09/ICCC_2021_paper_37.pdf; Simon Colton et al., *Generative Search Engines: Initial Experiments*, PROC. OF 12TH INT'L CONF. ON COMPUTATIONAL CREATIVITY (2021), https://computationalcreativity.net/iccc21/wpcontent/uploads/2021/09/ICCC_2021_paper_50.pdf; Ahmed Elgammal et al., *CAN: Creative Adversarial Networks Generating 'Art' by Learning About Styles and Deviating from Style Norms*, PROC. OF THE 8TH INT'L CONF. ON COMPUTATIONAL CREATIVITY (2017), https://computationalcreativity.net/iccc2017/iccc17_proceedings.pdf.

[22] "Deep learning" is a type of machine learning based on artificial neural networks in which multiple layers of computer processing are used to extract progressively higher-level features from data. Frank Emmert-Strieb et al., *An Introductory Review of Deep Learning for Prediction Models with Big Data*, 3 FRONTIERS A.I. 1, 1–2 (2020), https://www.frontiersin.org/articles/10.3389/frai.2020.00004/full.

[23] *See*, *e.g.*, Leo Gao et al., *The Pile: An 800GB Data Set of Diverse Text for Language Modeling*, ARXIV:2101.00027 at 1–6 (Dec. 31, 2020), https://arxiv.org/abs/2101.00027 (describing a large dataset aimed at training large-scale language models).

[24] *See*, *e.g.*, Giorgio Franceschelli & Mirco Musolesi, *Creativity and Machine Learning: A Survey*, ARXIV:2014.02726 (July 5, 2022),

replicated style, may consist of text, images, audio, or video. The speedy development of GenAI has shocked the public because of how well it fares on creative tasks like writing poetry and drawing images, and how well it can create synthesized content mimicking the work of real people.

Another big change has been the remarkable fluency with language that current AI models demonstrate. Only four years ago, large language models ("LLMs") would routinely "forget" basic parts of the conversations they were having with human partners or would incomprehensibly babble in the middle of answering a question. Now, these models are so adept with language that they can comfortably produce sentences indistinguishable from their human-authored counterparts and can "recall" earlier parts of a conversation with ease.

The first GenAI approaches that were introduced involved text-to-text, that is, a user input a textual question or instruction, and the AI returned a textual, often narrative, response by predicting the words in a sentence. There have been many such LLMs offered by Silicon Valley tech companies, including Google's Language Model for Dialogue Applications ("LaMDA" or "Bard"),[25] Meta's Large Language Model Meta AI ("LLaMA"),[26] and perhaps the most well-known of all, Open AI's Generative Pre-trained Transformer ("GPT") series.[27]

While ChatGPT may only have leapt into the general

---

https://arxiv.org/abs/2104.02726 (providing an overview of the history of computational creativity theories); Ian J. Goodfellow et al., *Generative Adversarial Networks*, ARXIV:1406.2661 (June 10, 2014), https://arxiv.org/abs/1406.2661 (proposing a new framework for using an adversarial process for estimating generative models).

[25] Eli Collins & Zoubin Ghahramani, *LaMDA: Our Breakthrough Conversation Technology*, THE KEYWORD BLOG (May 18, 2021), https://blog.google/technology/ai/lamda/.

[26] *Introducing LLaMA: A Foundational 65-Billion Parameter Large Language Model*, META AI BLOG (Feb. 24, 2023), https://ai.facebook.com/blog/large-language-model-llama-meta-ai/.

[27] *GPT-4 is OpenAI's Most Advanced System, Producing Safer and More Useful Responses*, OPENAI, https://openai.com/product/gpt-4 (last visited Nov. 10, 2023).

public's awareness following its release in late November, 2022,[28] significant advancements in the field of GenAI date back to as far as the 2010s. In 2014, the GenAI framework, Generative Adversarial Networks ("GANs"),[29] took a huge step forward in creating images, videos, and audio that appeared authentic. In this new framework, two networks "compete": A generative network drafts candidates and a discriminative network evaluates those candidates against true data to try to distinguish them. On the generative network's side, this leads to generated content that is more true-seeming. On the discriminative network's side, this leads to new findings about the characteristics that improve accuracy in matching the training data.

In 2017, Google introduced the transformer architecture,[30] which was another breakthrough in computer processing of natural language. Transformers do not require pre-labelled training data and can be trained in parallel, allowing much faster training than previous AI architectures. Many now well-known models, like the GPT series, are built using transformers, and each of the new GPT models is trained on progressively more data and can more accurately model human language than its predecessor(s). Another important change that began with GPT-3 was the introduction of reinforcement learning[31]—a process which includes human feedback—used to improve the output of an AI model. For LLMs, the addition of reinforcement learning allowed OpenAI, the creator of the GPT models, to endeavor to avoid having its models produce improper or offensive outputs.

ChatGPT—the model that took the Internet by storm—interacts with users in a dialogue fashion and is built on top of GPT-3.5. Because of its ability to understand user input, it can keep a natural flow of conversation, answering follow-up questions and responding to feedback along the way. ChatGPT amazed people in

---

[28] *Introducing ChatGPT*, OPENAI (Nov. 30, 2022), https://openai.com/blog/chatgpt.

[29] Goodfellow et al., *supra* note 24, at 1.

[30] *See* Ashish Vaswani et al., *Attention is All You Need*, ARXIV:1706.03762 [cs.CL] (Dec. 6, 2017), https://arxiv.org/abs/1706.03762 (proposing a new network architecture).

[31] *See generally*, REINFORCEMENT LEARNING: STATE-OF-THE-ART (Marco Wiering & Martin Otterlo eds., 2012) (covering major recent developments in reinforcement learning).

at least two ways. First, it could generate creative artifacts that appeared to be every bit as creative as those generated by humans, and that in some instances, could not be distinguished from the work of the artist it was replicating. And second, in conversations with humans, it could, in some circumstances, also convince its human conversational partner that they were conversing with another human. In doing so, it appeared to pass the Turing Test,[32] even going as far as to convince some people that it was sentient.[33]

ChatGPT can write poems in the style of Shakespeare and song lyrics in the style of Justin Bieber, all within a matter of seconds. Nonetheless, there are still many limitations to ChatGPT. Although it is designed to acknowledge its shortcomings rather than spout misleading or biased information, sometimes it still confidently answers questions like "Which is heavier, 1kg of feather or 1kg of

---

[32] The "Turing test," first described by Alan Turing in 1950, asks a human to determine which of two conversational partners is a human and which is a computational agent; an agent satisfies the test if it can confuse its conversational partner into thinking it is human. *See* A. M. Turing, *Computational Machinery and Intelligence*, 59 MIND 433 (1950) (describing the Turing test). Turing, himself, referred to his idea as the "imitation game," however others since then have reserved that moniker for one particular version of the test. *See Turing Test*, WIKIPEDIA https://en.wikipedia.org/wiki/Turing_test (last visited Nov. 10, 2023). The Turing test is widely considered the most influential test for intelligence in computers, although it not without criticism. *See id.*; *see also*, *e.g.*, Alison Pease & Simon Colton, *On Impact and Evaluation in Computational Creativity: A Discussion of the Turning Test and an Alternative Proposal*, PROCEEDINGS OF AISB '1: COMPUTING AND PHILOSOPHY (2011), https://discovery.dundee.ac.uk/en/publications/on-impact-and-evaluation-in-computational-creativity-a-discussion (criticizing the Turing test). If you would like to try your hand at chatting for two minutes and trying to figure out whether your conversational partner is a fellow human or a chatbot, *see Human or Not? A Social Turing Game*, AI21LABS, https://www.humanornot.ai/ (last visited Sep. 17, 2023) (allowing users to chat for two minutes and try to determine whether they just conversed with a human or an AI bot).

[33] *Google Fires Software Engineer Who Claims AI Chatbot Is Sentient*, THE GUARDIAN (July 23, 2022, 4:12 AM), https://www.theguardian.com/technology/2022/jul/23/google-fires-software-engineer-who-claims-ai-chatbot-is-sentient; *see also* Matt Meuse, *Bots Like ChatGPT Aren't Sentient. Why Do We Insist on Making Them Seem Like They Are?*, CBC RADIO (Mar. 17, 2023, 12:59 PM EST), https://www.cbc.ca/radio/spark/bots-like-chatgpt-aren-t-sentient-why-do-we-insist-on-making-them-seem-like-they-are-1.6761709 (criticizing the tendency of many to try to argue that AIs are sentient).

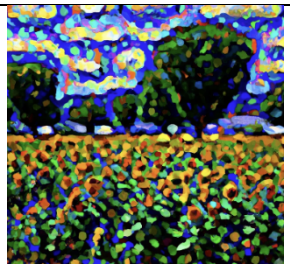
iron?" by incorrectly insisting that 1kg of iron is heavier. (It is obvious to most humans that since both are 1kg, their weight is the same, even though, in general, iron is more dense than feathers!) Chat GPT can also miss biases inherent in its own responses to leading questions or invent citations and references to publications or authors that do not exist. These faulty responses are often referred to as "hallucinations."[34]

Another example of models that use GPT-3 is DALL-E 2,[35] a deep learning model that can respond to specific textual prompts by producing responsive images. However, while DALL-E 2 can generate images from prompts like "Draw an illustration of a baby daikon radish in a tutu walking a dog," whether it has an actual understanding of the language in the prompt is questionable. It has limitations in dealing with negation and in making inferences using common sense. For instance, the following images generated by DALL-E 2 show how irrelevant or meaningless the images can be in response to open-ended prompts that require actual comprehension of the instruction, or where DALL-E 2 has insufficient image reference data associated with a complex, abstract concept included in a prompt.

---

[34] *See* Ziwei Ji et al., *supra* note 11 (explaining the tendency of language models to hallucinate false information).

[35] *See* Aditya Ramesh et al., *Hierarchical Text-Conditional Image Generation with CLIP Latents*, ARXIV:2204.01625 [cs.CV] (Apr. 13, 2022), https://arxiv.org/abs/2204.06125 (comparing a model with Dall-E).

 **Draw admissible evidence**	 **Draw admissible evidence in the style of Van Gogh**	 **Draw admissible evidence in the style of Picasso**
 **Draw inadmissible evidence**	 **Draw inadmissible evidence in the style of Van Gogh**	 **Draw inadmissible evidence in the style of Picasso**

On the other hand, VALL-E, a model for text-to-speech ("TTS") synthesis, focuses on the task of generating audio from a given text prompt as well as a "ground truth," an audio of the intended speaker that is at least three seconds in length.[36] Previously, TTS required clean data from a recording studio to produce output, meaning a lot of available data could not be used for training. This is no longer the case because VALL-E now accepts a wide variety of training data and leverages it to make better generalizations. To the naked ear, the generated audio is indistinguishable from the original speaker because VALL-E accounts for background noise in addition to just matching the speaker's voice.

---

[36] *See* Chengyi Wang et al., *Neural Codec Language Models Are Zero-Shot Text to Speech Synthesizers*, ARXIV:2301.02111 [cs.CL] (Jan. 5, 2023), https://arxiv.org/abs/2301.02111 (introducing the Vall-E language model and noting that a three-second recording of the target speaker is sufficient for an acoustic prompt).

All of these are merely examples of what can currently be done with GenAI. Open AI claims that GPT-4, which was released on March 14, 2023, is 19 percentage points better at producing factual responses than its predecessor.[37] Nonetheless, there is a lack of clarity of how GPT-4 was trained and the data set on which it was trained. It can generate complex computer code and can also directly identify properties of input images. While ChatGPT scored at the tenth percentile on the U.S. bar exam, GPT-4 passed it easily, scoring at the 90th percentile.[38]

Moreover, with the release of ChatGPT plugins on March 23, 2023,[39] ChatGPT is no longer limited to outdated information. It can now interact with real-time data to perform tasks in conjunction with other tools, like booking a trip using Expedia, or purchasing items on Instacart. Still, we are nowhere near the end of the development of these tools.[40] Not only can we expect GenAI to

---

[37] *See* OpenAI, *GPT-4 Technical Report*, arXiv.2303.08774 [cs.CL] 10 (Mar. 27, 2023), https://arxiv.org/abs/2303.08774 (reporting on the development of GPT-4).

[38] *Compare* Stephanie Wilkins, *How GPT-4 Mastered the Entire Bar Exam, and Why That Matters*, LAW.COM (Mar. 17, 2023, 11:35 AM EST), https://www.law.com/legaltechnews/2023/03/17/how-gpt-4-mastered-the-entire-bar-exam-and-why-that-matters/?kw=How%20GPT-4%20Mastered%20the%20Entire%20Bar%20Exam%2C%20and%20Why%20That%20Matters (finding GPT-4 scored at the 90th percentile on the bar), *with* Karen Sloan, *U.S. Bar Exam Pass Rate Drops for First-Time Takers*, REUTERS (Feb. 28, 2023, 2:13 PM EST), https://www.reuters.com/legal/legalindustry/us-bar-exam-pass-rate-drops-first-time-takers-2023-02-27/ ("Just over 78% of U.S. law school graduates who took the bar exam for the first time in 2022," passed, which was "down slightly from the 80% first-time pass rate in 2021 and represents a 6 percent decline from 2020's first-time pass rate of 84%."), *and* Alexander Overton, *Time for an End to the Bar Exams for Canadian Lawyers*, CANADIANLAWYER (May 14, 2021), https://www.canadianlawyermag.com/news/opinion/time-for-an-end-to-the-bar-exams-for-canadian-lawyers/356144 (Finding that, in Ontario, Canada—where three of the authors of this paper reside—"the bar exams pass rate is [already] north of 90 per cent. . . .").

[39] *ChatGPT plugins*, OPENAI, https://openai.com/blog/chatgpt-plugins (last visited Sept. 15, 2023).

[40] Luke Larson, *GPT-5: Release Date, Claims of AGI, Pushback, and More*, DIGITALTRENDS (Apr. 14, 2023), https://www.digitaltrends.com/computing/gpt-5-rumors-news-release-date/ ("OpenAI has officially stated that GPT-4.5 will be introduced in 'September or October 2023' as an 'intermediate version between GPT-4 and the upcoming GPT-5.'").

get better at what it does, it will also be able to take on increasingly complex tasks with lesser degrees of human involvement.

III. SOME ISSUES FOR JUDGES TO PONDER

A. Do We Need New Rules of Evidence to Address GenAI?

When cases such as those described in the hypotheticals above reach the courts—and they will with alarming speed—judges will be called upon to make determinations about the authenticity and admissibility of evidence that may be produced by GenAI applications, or evidence that may be truly human-generated or of unknown origin, but challenged as deepfake. There is no question that proffering, challenging, and ruling on digital evidence just got harder.

In the main, the existing Federal Rules of Evidence and their state counterparts are written to provide general guidance to trial judges and attorneys in a vast array of cases, and only occasionally do they provide rules geared specifically to any particular type of technical evidence. This is because revising the Federal Rules of Evidence and their state counterparts is a time-consuming process, while technology in general—and GenAI in particular—change at a breakneck pace.[41] Although there has been a recent call to amend the Federal Rules of Evidence to eliminate the role of the jury in determining the authenticity of digital and audiovisual evidence in response to the appearance of deepfakes,[42] such a change would involve a substantial departure from the current evidentiary framework, and would take considerable time to adopt, making it infeasible as a practical solution. We simply cannot change the rules of evidence with the introduction of each new technological development. Meanwhile, cases involving evidence known to be the product of GenAI applications, and evidence of unknown or challenged origin, but potentially AI-generated—*e.g.*, deepfake evidence—will reach the courts sooner than we know it. Judges and

---

[41] *See* Paul W. Grimm, Maura R. Grossman & Gordon V. Cormack, *Artificial Intelligence as Evidence*, 19 NW. J. TECH. & INTELL. PROP. 9, 84 (2021) (hereinafter "Grimm, Grossman & Cormack").

[42] Rebecca A. Delfino, *Deepfakes on Trial: A Call to Expand the Trial Judge's Gatekeeping Role to Protect Legal Proceedings from Technological Fakery*, 74 HASTINGS L.J. 293, 297 (2023).

attorneys will undoubtedly be required to address this evidence under the current rules of evidence.

Under the existing Federal Rules of Evidence, the key issues that must be addressed in determining the admissibility of GenAI evidence—as with any evidence—are: (i) relevance (Fed. R. Evid. 401), (ii) authenticity (Fed. R. Evid. 901 and 902), (iii) the judge's role as an evidentiary gatekeeper (Fed. R. Evid. 104(a)), (iv) the jury's role as a decider of contested facts relating to the authenticity of evidence (Fed. R. Evid. 104(b)), and (v) the need to exclude evidence that, while relevant, is unfairly prejudicial (Fed. R. Evid. 403).

Judges need to bear in mind that the Rules of Evidence were intended to be applied flexibly, "to promote the development of evidence law,"[43] meaning that the existing rules should not be rigidly applied in the face of technological advancements. Instead, they should be adapted to permit their application to new technologies and the evidentiary challenges that accompany them, such as those now posed by GenAI and deepfake evidence.[44] If this approach is to be followed, then in addition to the Fed. R. Evid. cited above, judges must also be informed by the rule that requires them to be the gatekeepers determining the admissibility of scientific, technical, and specialized evidence (Fed. R. Evid. 702). This rule, in its prior version—and now in its recently amended version[45]—requires the

---

[43] Fed. R. Evid. 102.

[44] For a comprehensive analysis of these issues as they relate to AI evidence, *see* Grimm, Grossman & Cormack, *supra* note 41, at 84–105.

[45] The proposed changes to Fed. R. Evid. 702 which took effect on December 1, 2023, are subtle, but very significant. The amendment adds the language "[if] the proponent demonstrates to the court that it is more likely than not that" the proposed expert's scientific, technical, or specialized knowledge will help the finder of fact to understand the evidence or decide a fact that is in issue, the expert's testimony is based on sufficient facts or data, the expert's testimony is the product of reliable principles and methods, and that the "expert's opinion reflects a reliable application of" the principles and methods to the fact of the case. *Proposed Amendments to the Fed. R. Evid.[], Rule 702 (Testimony by Expert Witness)*, Advisory Comm. on Evid. Rules, Memorandum to the Standing Comm. (May 15, 2022), in Comm. on Rules of Prac. & Proc., Agenda Book, Appendix A: Rules for Final Approval, at 891–96 (June 7, 2022), https://www.uscourts.gov/sites/default/files/2022-06_standing_committee_agenda_book_final.pdf. The new rule clarifies that the proponent of the expert evidence has the burden of demonstrating its helpfulness,

trial judge to ensure that scientific and technical evidence that is beyond the ability of lay juries to understand without expert assistance, but will be helpful to the jury in deciding the issues they must resolve, is based on sufficient facts and supported by reliable methodology which has been reliably applied to the facts of the particular case.[46] In determining whether the methodology or principles that underly the scientific or technical evidence are "reliable,"[47] judges must ensure that the evidence is both *valid* (*i.e.*, accurately measures or reflects what it is supposed to measure or reflect) and *reliable* (*i.e.*, is consistently accurate when applied under substantially similar facts and circumstances). Finally, but perhaps most importantly, when evaluating the admissibility of evidence of disputed origin that potentially is GenAI or deepfake evidence, trial judges must pay particular attention to the need to avoid the unfair prejudice that can occur if insufficiently valid and reliable evidence is allowed to be presented to the jury. Thus, Fed. R. Evid. 403 is particularly important in assessing the authenticity of potential GenAI or deepfake evidence. We outline below the steps that judges should follow when faced with determining the admissibility of such evidence.

B. What's a Judge to Do? New Wine in Old Bottles!

As a preliminary matter, when exercising their gatekeeping function to rule on challenged evidence that is being offered as

---

factual sufficiency, reliable basis, and reliable application to the facts of the case by a "preponderance" of evidence (*i.e.*, more likely than not). In addition, it underscores the obligation of the trial court to determine (under Fed. R. Evid. 104(a)), as a condition of admissibility of the scientific, technical, or specialized evidence, that the proponent has met its burden before the fact finder is allowed to consider the evidence in the first place. In this regard, the Advisory Committee's Note to the proposed rule change reflects the view of the Evidence Rules Advisory Committee that federal judges had not adequately been fulfilling this preliminary screening role under Fed. R. Evid. 702. *See id.*, Committee Note at 892–93.

[46] *See* Grimm, Grossman & Cormack, *supra* note 41, at 95-97.

[47] The rules of evidence conflate two distinct but related concepts—validity and reliability—under the single umbrella term "reliability." Technical evidence has *validity* if it accurately does what it was designed to do; it has *reliability* if it consistently is accurate when applied to the same or substantially similar circumstances. AI evidence needs to have both validity and reliability. *See* Grimm, Grossman & Cormack, *supra* note 41, at 48.

"authentic," but which, in fact, could be GenAI evidence—deepfakes being the most common example—as well as evidence that is acknowledged to be GenAI, judges should use Fed. R Evid. 702 and the *Daubert* factors[48] to evaluate the validity and reliability of the challenged evidence and then make a careful assessment of the unfair prejudice that can accompany introduction of inaccurate or unreliable technical evidence. Under such an approach, a showing that evidence is merely more likely than not what it purports to be (*i.e.*, the standard of mere preponderance) should not be determinative of admissibility. The court must also consider the potential risk, negative impact, or untoward consequences that could occur if the evidence turns out to be fake, or insufficiently valid and reliable. In other words, when the risk of an unfair or erroneous outcome is high, and the evidence of authenticity is low, the evidence should be excluded. Judges who follow the following steps will be in the best position to make these important determinations.

1. STEP 1: Scheduling Order.

When issuing a scheduling order in a civil or criminal case, the court should set a deadline requiring a party that intends to introduce evidence that is or could potentially be based on a GenAI application to disclose the nature of that evidence to the opposing party and the court sufficiently in advance of trial or a hearing to permit opposing counsel to determine whether they intend to challenge the admissibility of that evidence, and whether they intend to seek

[48] The *Daubert* Factors were added to the Fed. R. Evid. in 2000, following the U.S. Supreme Court's decisions in *Daubert* v. *Merrell Dow Pharm., Inc.*, 509 U.S. 579 (1993) and *Kumho Tire Co.* v. *Carmichael*, 119 S. Ct. 1167 (1999). While Fed. R. Evid. 702 was not meant to codify the *Daubert* decision, the factors discussed therein relating to the determination of the reliability of scientific or technical evidence are instructive in determining whether Fed. R. Evid. 702's reliability requirement has been met. The *Daubert* Factors are: "(1) whether the expert's technique or theory can be or has been tested . . .; (2) whether the technique or theory has been subject to peer review and publication; (3) the known or potential rate of error of the technique or theory when applied; (4) the existence and maintenance of standards and controls; and (5) whether the technique or theory has been generally accepted in the scientific [or technical] community." Advisory Committee Note, Fed. R. Evid. 702 (2000). For further discussion on the usefulness of the *Daubert* factors in determining whether to admit AI Evidence, *see* Grimm, Grossman & Cormack, *supra* note 41, at 95-97.

discovery in order to mount a challenge to such evidence. Similarly, the scheduling order should include a deadline for the party against whom the actual or potential GenAI evidence will be introduced to advise the proponent of that evidence, and the court, of its intent to challenge the evidence and to request discovery in order to challenge its admissibility.

When discovery is sought but is opposed by the proponent of the challenged evidence, the court should hold a hearing (which may be informal or formal, as needed) to determine what discovery is requested, the objections to that discovery, and to issue an order outlining the discovery (if any) that will be permitted. If ordering discovery, the court should consider issuing a protective order to protect confidential trade secrets relating to any applicable AI system, algorithm, or data, if requested to do so. The scheduling order should set a deadline for the completion of the discovery and deadlines for the party intending to challenge the proffered evidence as AI-generated or deepfake to file a motion challenging the evidence, as well as the proponent's opposition to the motion to exclude, and the moving party's reply.

A slightly different approach is necessary in those cases where a party is offering evidence that it does not acknowledge to be the product of a GenAI application (*i.e.*, evidence that the non-offering party may allege to be deepfake evidence but the offering party believes is human-generated or genuine). In such cases, the offering party will not meet the deadline in the scheduling order for disclosure of GenAI evidence because it contends that the evidence is not the product of such technology. Nonetheless, the pretrial order will include a deadline for disclosure of witnesses and other evidence the parties intend to introduce, and the potential deepfake evidence will have been subject to discovery under Fed. R. Civ. P. 34 or Fed. R. Crim. P. 16(a)(1)(E) and 16(b)(1)(A). The party that contends that evidence that has been disclosed and/or produced during discovery is, in fact, a deepfake would then be able to request a conference with the court pursuant to Fed. R. Civ. P. 16 or Fed. R. Crim. P. 16.1 to request discovery in order to challenge the possible deepfake evidence, and the court would then proceed as set forth above for cases where a party acknowledges that it intends to introduce GenAI evidence.

2. Step 2: The Hearing.

When a challenge is made to the introduction of evidence as AI-generated or deepfake, the court should set an evidentiary hearing to develop the facts necessary to rule on the admissibility of the challenged evidence. Because the outcome of this ruling may have a substantial effect on whether there will be a trial, the hearing should be scheduled far enough in advance of trial for the evidentiary record to be made and evaluated by the judge, and for a ruling made on the admissibility of the challenged evidence. These hearings can be involved and the court should schedule enough time to ensure that the record is sufficiently complete. At the hearing, the proponent must meet their burden of establishing the relevance of the evidence (under Fed R. Evid. 401) and its authenticity by at least a preponderance of the evidence (under Fed. R. Evid. Rules 901 and 902). The opposing party should have the opportunity to introduce evidence challenging the relevance and authenticity of the proffered evidence, especially with respect to its validity and reliability, including any challenges to the methodology or principles underlying the data, training, or development of the AI system that generated the evidence. The proponent of the evidence should have the opportunity to rebut this evidence. Finally, the court should require the proponent of the evidence and the opposing party to address the potential risk of unfair or excessive prejudice that could result from introducing the proffered evidence—particularly if it should turn out to be invalid, unreliable, or a deepfake—based on the evidentiary record established at the trial.

3. Step 3: The Ruling.

Following the hearing, the court should carefully consider the evidence introduced and arguments made at the hearing and issue a ruling. In so doing, the court must assess whether the proponent of the evidence sufficiently met its burden of authenticating the evidence. The ruling should address the relevance, authentication, and prejudice arguments, and the court should pay particular attention to its conclusions regarding the validity and reliability of the challenged evidence and weigh the relevance of the proffered evidence against the risk of an unfair or excessively prejudicial outcome. Where the evidence may be highly prejudicial, a mere

preponderance may very well be insufficient. The judge should take full advantage of the analytical factors found in Fed. R. Evid. 702 and the *Daubert* factors in assessing the validity and reliability of the evidence.

On the question of authenticity, if the court determines that the facts are such that a reasonable jury could find that the challenged evidence more likely than not is authentic, but that a reasonable jury also could find that the challenged evidence more likely than not is not authentic, then this presents an issue of conditional relevance under Fed. R. Evid. 104(b). The rule requires the disputed facts regarding authenticity to be presented to the jury for its ultimate determination of authentication,[49] but only if the judge rules that,

[49] Fed. R. Evid. 104(b) deals with circumstances in which the relevance of proffered evidence depends upon the existence of a particular fact or facts, a situation sometimes referred to as "conditional relevance." *See* Advisory Committee Note to Fed. R. Evid. 104(b) (1972). Rule 104(b) itself provides that "[w]hen the relevance of evidence depends on whether a fact exists, proof must be introduced sufficient to support a finding that the fact does exist. The court may admit the proposed evidence on the condition that the proof be introduced later." Rule 104(b) must be considered in concert with Fed. R. Evid. 104(a), which states that "[t]he court must decide any preliminary question about whether . . . evidence is admissible." These two rules allocate the responsibility for determining the admissibility of evidence between the trial judge and the jury, when the underlying facts that establish the relevance of proffered evidence are challenged. The Advisory Committee Note to Rule 104(b) helpfully discusses this allocation of responsibility as follows: "If preliminary questions of conditional relevancy were determined solely by the judge, as provided by subdivision (a), the functioning of the jury as a trier of fact would be greatly restricted and, in some cases, virtually destroyed. These are appropriate questions for juries. Accepted treatment, as provided in the rule, is consistent with that given fact questions generally. The judge makes a preliminary determination whether the foundation evidence is sufficient to support a finding of fulfillment of the condition. If so, the item is admitted. If, after all the evidence on the issue is in, pro and con, the jury could reasonably conclude that fulfilment of the condition is not established, the issue is for them. If the evidence is not such as to allow a finding, the judge withdraws the matter from their consideration." In the context of evidence that is challenged as deepfake, the judge must initially assess whether the proponent has proffered sufficient facts that the challenged evidence is authentic, namely that the party introducing the evidence has shown, more likely than not, that it is what they claim it is. If the judge concludes that this threshold has not been established, the evidence is excluded. However, if the judge decides that this threshold has been established, the evidence is admitted for the jury to consider, but the opposing party may introduce evidence to rebut the proponent's authenticity evidence. If, after considering the proponent's and the opponent's

based on the hearing, there is not unfair or excessive prejudice to the opposing party in allowing the jury to consider the evidence, given the relevance of the disputed evidence, and the potential for an erroneous or unfair outcome if the jury considers it. If the judge determines that allowing the jury to decide the disputed authenticity of the evidence raises too great a risk of unfair or excessive prejudice to the party against whom the evidence is being offered, the judge should exclude it, exercising its authority under Fed. R. Evid. 104(a) to be the gatekeeper of what the jury is allowed to consider.

The changes to Fed. R. Evid 702, which took effect on December 1, 2023, make clear that highly technical evidence, such as that involving GenAI and deepfakes, create an enhanced need for trial judges to fulfill their obligation to serve as gatekeepers under Fed. R. Evid. 104(a), to ensure that only sufficiently authentic, valid, reliable—and not unfairly or excessively prejudicial—technical evidence is admitted. This role requires the judge to hold the proponent of the evidence to its obligation to meet the foundational requirements of Fed. R. Evid. 401, 901, and 702. This is especially so because, with the proliferation of deepfake evidence and the increased public awareness of it, courts must keep in mind that the cost of failing to fulfill their gatekeeping role may result in juries believing inauthentic deepfake evidence, or, conversely disbelieving

---

evidence, the jury concludes that the evidence is not authentic (*i.e.*, it is a deepfake), then the judge instructs the jury to disregard it and not to consider it in reaching their verdict. Fair enough in the abstract, but the jury will already have been exposed to the deepfake evidence, and—as we will explain (*infra* p. 28 & notes 55, 56)—it may not be so easily disregarded when the jury deliberates. As the saying goes, you cannot "unring a bell." It is our position that when judges undertake their Fed. R. Evid. 104(a) preliminary evaluation of whether the jury may hear evidence that is challenged as a deepfake, they also should consider the evidence proffered by the party opposing the evidence as to why it contends that it is fake, and then employ Fed. R. Evid. 403 to assess whether allowing the jury to consider the potential deepfake evidence under Fed. R. Evid 104(b) would expose the opposing party to unfair or excessive prejudice. If it would, then the judge should not allow the potential deepfake to be presented to the jury. In making this determination, the judge should evaluate the importance of the potential deepfake evidence when considered in light of all the other evidence that has been or will be admitted. If the potential deepfake evidence is corroborated by other evidence that is admissible, then the danger of unfair or excessive prejudice is considerably lessened. But if the potential deepfake is the only evidence offered to prove a fact that is critical to the resolution of the dispute, then the danger of unfair or excessive prejudice is great.

authentic evidence, because it has been wrongly characterized as deepfake by the party against whom it has been introduced. Either circumstance undermines accurate factfinding and fair trial outcomes.

While the focus of this article thus far has been on evidentiary issues, GenAI can be expected to raise additional questions for the court. We will briefly touch on a few of them.

C. Will Every Case Now Require an GenAI Expert?

The aforementioned increase in evidentiary hurdles that will be imposed on both the proponent of actual or suspected GenAI or deepfake evidence, as well as the challenger of such evidence, can be expected to require—at least for the immediate future—a greater need for technical and forensic experts who are well versed in GenAI and deepfakes. This will obviously serve to increase the cost of litigation in an already unaffordable justice system. These hurdles can be expected to cause a crisis for criminal defendants and public defenders who simply cannot afford the kinds of expensive experts that will be needed to mount a proper defense. More appeals based on a claim of ineffective assistance of counsel may also result. Right now, the technology available is insufficiently accurate or reliable to detect AI-generated or deepfake content; even OpenAI admits that its detector should not be used as a primary decision-making tool.[50]

We are already locked in an intractable arms race where adversarial attacks are proliferating at the same, if not greater, speed than secure solutions; in fact, at present, the development of better GenAI detectors may actually contribute to the development of GenAI that is harder to detect. This is because, as explained above,[51]

---

[50] *See* Kirchner et al., *supra* note 9 ("**Our classifier is not fully reliable.** In our evaluations on a 'challenge set' of English tests, our classifier correctly identifies 26% of AI-written (true positives) as 'likely AI-written,' while incorrectly labeling human-written text as AI-written 9% of the time (false positives)." (emphasis in original)). *See also* Ann-Marie Alcántara, *AI-Created Images Are So Good Even AI Has Trouble Spotting Some*, THE WALL STREET JOURNAL (Apr. 11, 2023, 8:00 AM ET), https://www.wsj.com/articles/ai-created-images-are-so-good-even-ai-has-trouble-spotting-some-8536e52c?mod=e2twd.

[51] *See supra* p. 11 & note 30.

one approach for advancing GenAI uses GAN networks, and better detection algorithms also mean better training material for GenAI. So, it is not just an arms race, it is a permanent deadlock.

While an extended discussion of the role of experts in this new GenAI world is beyond the scope of this paper, it is worth noting that if the parties' experts do not provide the judge with sufficient information concerning the validity, reliability, or prejudice factors to allow the judge to rule, the judge can appoint an expert (under Fed. R. of Evid. 706) or a technical advisor (under its inherent authority), to educate the court on the GenAI or other technology at issue.[52]

D. Will Juries Still Be Able to Do Their Jobs?

GenAI and deepfake evidence can also be expected to throw a monkey wrench in the role of juries tasked with determining the proper weight to give evidence admitted from black-box AI systems that they little understand, and to audio, video, and documentary evidence that they can no longer assess or trust using their own senses. Research has already demonstrated that humans are unable to reliably distinguish AI-generated faces from real faces in photographs and find the AI-generated faces to be more trustworthy.[53] Audiovisual evidence is particularly scary. Studies have shown that "jurors who hear oral testimony along with video testimony are **650%** more likely to retain the information," and that "video evidence powerfully affects human memory and perception

---

[52] *See* Fed. R. Evid. 706 (providing for court-appointed expert witnesses); *see generally, e.g.*, Robert L. Hess II, *Judges Cooperating with Scientists: A Proposal for More Effective Limits on the Federal Judge's Inherent Power to Appoint Technical Advisors*, 54 VAND. L. REV. 547 (2001), https://scholarship.law.vanderbilt.edu/vlr/vol54/iss2/8/; Samuel H. Jackson, *Technical Advisors Deserve Equal Billing With Court Appointed Experts in Novel And Complex Scientific Cases: Does The Federal Judicial Center Agree?*, 28 ENV'TL. L. 431 (1998), https://www.jstor.org/stable/43266661.

[53] *See* Sophie J. Nightingale and Hany Farid, *AI-synthesized faces are indistinguishable from real faces and more trustworthy*, 119 PNAS 8, 1-3 (2022), https://www.pnas.org/doi/10.1073/pnas.2120481119; *see also* Zeyu Lu et al., *Seeing is not always believing: A Quantitative Study on Human Perception of AI-Generated Images*, ARXIV:2304.13023 [cs.AI] (Apr. 25, 2023), https://arxiv.org/abs/2304.13023 (showing that humans cannot distinguish between real photos and AI-created fake photos to a significant degree).

of reality."[54] Thus, even when people are aware that audiovisual evidence might be fake, it can still have an undue impact on them because they align their perceptions and memories to coincide with what they saw and heard on the recording in spite of their skepticism.[55]

Moreover, because the evidence placed before them now has a real likelihood of deceiving them, jurors are also more inclined to suspect the veracity of genuine evidence—a consequence of "truth decay"[56]—leading to cynicism and decision-making that may be based on conscious or unconscious biases, stereotypes, affective responses to the parties or their counsel, and other unknown and uncontrolled factors.

In a recent law review paper that we referenced earlier, Loyola Law School Professor Rebecca Delfino expressed concern about the emergence of "the deepfake defense,"[57] which Bobby Chesney and Danielle Citron had previously termed "the liar's dividend," in a prescient 2019 paper.[58] Essentially, the idea is that as people become more aware of how easy it is to manipulate audio

---

[54] Rebecca A. Delfina, *supra* note 42, at 311 & notes 101, 102 thereon (emphasis added).

[55] *See* Kimberly A. Wade et al., *Can Fabricated Evidence Induce False Eyewitness Testimony?*, 24 APPLIED COG. PSYCH. 899 (2010), https://onlinelibrary.wiley.com/doi/10.1002/acp.1607. This study showed the profound impact video can have on reconstructing personal observations. Sixty college students who were placed in a room to engage in a computerized gambling task were each later shown a digitally altered video depicting another subject cheating, when none had actually done so. Nearly half of the subjects were willing to testify that they had personally witnessed another subject cheating in real life after viewing the fake video. *See also* Hadley Liggett, *Fake Video Can Convince Witnesses To Give False Testimony*, WIRED (Sept. 14, 2009, 6:02 PM), https://www.wired.com/2009/09/falsetestimony/ (reporting on study).

[56] *See generally* Bobby Chesney & Danielle Keats Citron, *21st Century-Style Truth Decay: Deep Fakes and the Challenge for Privacy, Free Expression, and National Security*, 78 MD. L. REV. 882 (2019), https://digitalcommons.law.umaryland.edu/cgi/viewcontent.cgi?article=3834&context=mlr.

[57] *See* Rebecca Delfino, *supra* note 42, at 310-13.

[58] *See* Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. R. 1753, 1758 (2019), https://www.californialawreview.org/print/deep-fakes-a-looming-challenge-for-privacy-democracy-and-national-security ("[D]eep fakes make it easier for liars to avoid accountability for things that are in fact true.").

and visual evidence, defendants will use that skepticism to their benefit.[59] The "deepfake defense" has already been offered in several cases. In one case, lawyers for Elon Musk sought to argue that a YouTube video that had been posted online for seven years—which contained statements made by their client at a tech conference in 2016—could easily have been altered.[60] In another, two defendants on trial for their participation in the January 6th insurrection attempted to argue that videos showing them at the Capitol on that date could have been created or manipulated by AI.[61] In both cases, the court was not having any of it, but this issue poses a real threat to the justice system, particularly in criminal cases.

E. Is GenAI a Boon to Access to Justice or Does it Present a Whole New World of Opportunity for Bringing Vexatious Lawsuits?

Gen AI systems can now assist would-be litigants who lack legal representation—the vast majority of the parties in civil cases in state and local courts today,[62] and often individuals from racialized or otherwise marginalized communities—in identifying claims and in drafting complaints and other pleadings, and this is undoubtedly a welcome development. These individuals can now use GenAI to determine whether they satisfy the elements of various claims and generate customized language specific to individual circumstances and particular jurisdictions. But along with this potentially positive impact, malicious *pro se* filers also can now prepare simultaneous filings in courts around the country, permitting them to flood the courts with dozens of potentially duplicate, frivolous submissions. Their pleadings may even include citations to cases that do not exist. Apparently, "[d]ebt collection agencies are already flooding courts and ambushing ordinary people with thousands of low-quality, small-dollar cases. Courts are woefully

---

[59] Shannon Bond, *People are trying to claim real videos are deepfakes. The courts are not amused*, NPR (May 8, 2023, 5:01 AM EDT), https://www.npr.org/2023/05/08/1174132413/people-are-trying-to-claim-real-videos-are-deepfakes-the-courts-are-not-amused.

[60] *See id.*

[61] *See id.*

[62] *See* Anna E. Carpenter et al., *America's Lawyerless Courts*, ABA LAW PRACTICE MAGAZINE (Jul. 18, 2022), https://www.americanbar.org/groups/law_practice/publications/law_practice_magazine/2022/july-august/americas-lawyerless-courts/.

unprepared for a future where anyone with a chatbot can become a high-volume filer, or where ordinary people might rely on chatbots for desperately-needed legal advice."[63] The goal, in some of these cases, is to "[t]urn hard-to-collect debt into easy-to-collect wage garnishments. . . . The easiest way for that to happen? When the defendant doesn't show up, defaulting the case. . . . When a case does default, many courts will simply grant whatever judgment the plaintiff has requested, without checking whether the plaintiff has provided adequate (or any) documentation that the plaintiff owns the debt, that the defendant still owes the debt, or whether the defendant has been properly notified of the case."[64]

DoNotPay—an early self-help application that first appeared in 2015 to help fight parking tickets, and that, until mid-2023, touted itself as "The World's First Robot Lawyer," which can "sue anyone at the press of a button"[65]—recently found itself in hot water when a Chicago law firm brought a putative class suit against the company in San Francisco state court for practicing law without a license and violating California's unfair competition law.[66] Regardless of whether one views GenAI as a genuine boon to access to justice,[67] or as a sharp instrument for bludgeoning one's opponents, the justice system is ill-equipped to manage a massive influx of new cases that may be chock full of defects, false affidavits, faulty notarizations, incomplete paperwork, inadequate documentation, and so on, and

---

[63] Keith Porcaro, *Robot Lawyers Are About to Flood the Courts*, WIRED (Apr. 13, 2023, 7:00 AM EDT), https://www.wired.com/story/generative-ai-courts-law-justice/.

[64] *Id.*

[65] The World's First Robot Lawyer, DoNotPay, https://donotpay.com/ [https://web.archive.org/web/20230601033641/https:/donotpay.com/] (last visited June 1, 2023) (DoNotPay has since changed its tagline to "Your AI Consumer Champion," but the original homepage can be viewed at the archived link).

[66] Sara Merken, *Lawsuit Pits Class Action Firm Against 'Robot Lawyer' DoNotPay*, Reuters (Mar. 9, 2023, 3:10 PM EST), https://www.reuters.com/legal/lawsuit-pits-class-action-firm-against-robot-lawyer-donotpay-2023-03-09/. The case has since been removed to federal district court in the Northern District of California. *See* Faridian v. DoNotPay Inc., No.3:2023-cv-01692 (N.D. Cal. Apr. 7, 2023).

[67] *See*, *e.g.*, Andrew T. Holt, *Legal AI-d to Your Service: Making Access to Justice a Reality*, JETLaw Blog (Feb. 4, 2023), https://www.vanderbilt.edu/jetlaw/2023/02/04/legal-ai-d-to-your-service-making-access-to-justice-a-reality/.

like science fiction magazine Clarkesworld discussed above,[68] may buckle under the weight of such submissions.

F. Will Substantive Intellectual Property Law Have to Change to Accommodate Gen AI?

GenAI can be expected to give rise to numerous novel questions involving substantive intellectual property ("IP") law, which we can only briefly mention in passing here.[69] The U.S. Copyright Office has repeatedly issued policy guidance stating that material generated by AI is not eligible for copyright protection, as the goal of copyright is to protect efforts engaged in by *humans*; since AI does not engage in creative labor, it cannot create copyrighted works.[70] The Copyright Office has distinguished, in particular, between works "produced by a machine or mere mechanical process that operates randomly or automatically without any creative input or intervention from a human author" and those created "by a human being."[71] However, as creators start to incorporate GenAI work product as a component of their creative processes, this straight-line separation may become increasingly hard to define.

A recent test case is illustrated by the copyright registration mess involving Kristina Kashtanova, who created a comic book, *Zarya of the Dawn*, using Midjourney as the GenAI art creator, and registered a copyright for the book, including the Gen-AI-created

---

[68] *See supra* pp. 3 & note 6.

[69] For more detailed discussions, *see*, *e.g.*, *A New Generation of Legal Issues Part 1: The Latest Chapter in Copyrightability of AI-Generated Works,* PERKINS COIE LLP, (Jan. 26, 2023), https://www.perkinscoie.com/en/news-insights/a-new-generation-of-legal-issues-part-1-the-latest-chapter-in-copyrightability-of-ai-generated-works.html; *A New Generation of Legal Issues Part 2: First Lawsuits Arrive Addressing Generative AI,* PERKINS COIE LLP, (Apr. 20, 2023), https://www.perkinscoie.com/en/news-insights/first-lawsuits-arrive-addressing-generative-ai.html.

[70] See *Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence*, 88.51 Fed. Reg. 16,191 (Mar. 16, 2023) (to be codified at 37 C.F.R. pt. 202) ("In the Office's view, it is well-established that copyright can protect only material that is the product of human creativity. Most fundamentally, the term 'author,' which is used in both the Constitution and the Copyright Act, excludes non-humans.").

[71] *Id.* at 16192.

images. The copyright, which was originally granted, was subsequently withdrawn and replaced by a copyright grant only for the comic book's text, as well as the selection, coordination, and arrangement of its written and visual elements.[72] "The images themselves, however, 'are not the product of human authorship,' and the registration originally granted for them has been canceled. To justify its decision, the Copyright Office cite[d] previous cases where people weren't able to copyright words or songs that listed 'non-human spiritual beings' or the Holy Spirit as the author—as well as the infamous incident where a selfie was taken by a monkey."[73] Meanwhile, the Copyright Office also has suggested that merely providing simple prompts to an AI system will not, without more, qualify the resultant work for any copyright protection.[74] What more may be required to qualify as sufficiently creative is unclear.

Another issue arises with respect to the existing copyrights of materials used for training GenAI systems. It is not clear whether training on a collection of art, music, or text qualifies as "fair use," particularly if it competes in the same market as the original work,[75]

---

[72] *See* Richard Lawler, *The US Copyright Office says you can't copyright Midjourney AI-generated images*, THE VERGE (Feb. 22, 2023, 9:06 PM EST), https://www.theverge.com/2023/2/22/23611278/midjourney-ai-copyright-office-kristina-kashtanova.

[73] *Id.* (quoting letter from Robert J. Kasunic, Assoc. Reg. of Copyrights and Dir. of the Off. of Registration Pol'y & Prac., U.S. Copyright Office., to Kris Kashtanova's Law., Van Lindberg, at 4 (Feb. 21, 2023), https://www.copyright.gov/docs/zarya-of-the-dawn.pdf). *See also* Sarah Jeong, *Appeals court blasts PETA for using selfie monkey as 'an unwitting pawn,'* THE VERGE (Apr. 24, 2018, 8:00 AM EDT), https://www.theverge.com/2018/4/24/17271410/monkey-selfie-naruto-slater-copyright-peta.

[74] *See* letter from Robert J. Kasunic, *supra* note 73, at 8-9. *See also Whose Copyright Is It Anyway? Copyright Office Stakes Out Position on Registration of AI-Generated Works*, PERKINS COIE LLP, (Mar. 21, 2023), https://www.perkinscoie.com/en/news-insights/whose-copyright-is-it-anyway-copyright-office-stakes-out-position-on-registration-of-ai-generated-works.html.

[75] *See*, *e.g.*, Mark A. Lemley and Bryan Casey, *Fair Learning*, SSRN (Feb. 14, 2020), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3528447; Michael W. Carroll, *Copyright and the Progress of Science: Why Text and Data Mining Is Lawful*, 53 UC DAVIS L. REV. 893 (2019), https://lawreview.law.ucdavis.edu/issues/53/2/articles/files/53-2_Carroll.pdf; Benjamin L.W. Sobel, *Artificial Intelligence's Fair Use Crisis*, 41 COLUM. J. OF L. & THE ARTS. 41 45-97 (2017). For two of the authors' take on the application

and the providers of several visual GenAI systems have already been sued by artists who are concerned that their own back catalogs are being used—without permission—to train models that compete with their own work.[76] Questions of compensation for copyright holders are clearly ripe for litigation, as is determining how copyright holders can opt out having their own materials be used as training data for GenAI models.

An additional concern is that the output of AI-generated art systems may infringe or dilute existing trademarks. For example, in response to a prompt, Midjourney might create a character that looks a little too much like Mickey Mouse or She-Ra, or one that uses the Nike swoosh symbol. In these circumstances, there are real questions about who (if anyone) might be liable for that, and what a take-down procedure might look like in the GenAI context.[77]

The outcome in the *Getty Images* case referenced above[78] may provide some guidance about whether the incorporation of a trademark in AI-generated output can constitute trademark infringement or give rise to a trademark dilution claim under 15 U.S.C. §1125(c). The *Getty Images* Complaint alleges that the images generated by Stability AI are infringing because they are likely to cause confusion among viewers. In particular, viewers might believe that the AI-generated images in some way suggest a business relationship between Stability AI and Getty Images that does not exist. Viewers might incorrectly believe that Getty Images had granted Stability AI the right to use its marks, or that Getty

---

of the fair-dealing exception in the Canadian Copyright Act in this context, *see* Dan Brown, Lauren Byl & Maura R. Grossman, *Are machine learning corpora 'fair dealing' under Canadian Law?*, UWSPACE (2021), https://uwspace.uwaterloo.ca/bitstream/handle/10012/17708/ICCC_2021_paper_68.pdf?sequence=1&isAllowed=y.

[76] *See* cases cited at *supra* note 15, and *in A New Generation of Legal Issues Part 2*, *supra* note 69. *See also*, *e.g.*, Thomas James, *Does AI Infringe Copyright?*, COKATO COPYRIGHT ATTORNEY: THE L. BLOG OF THOMAS JAMES (Jan. 24, 2023), https://thomasbjames.com/does-ai-infringe-copyright/; Blake Brittain, *Lawsuits accuse AI content creators of misusing copyrighted work*, REUTERS (Jan. 17, 2023, 3:05 PM EST), https://www.reuters.com/legal/transactional/lawsuits-accuse-ai-content-creators-misusing-copyrighted-work-2023-01-17/.

[77] *See What Does AI Mean for Trademarks?*, LICENSING INT'L, (Feb. 22, 2023), https://licensinginternational.org/news/what-does-ai-mean-for-trademarks/.

[78] *See supra* note 15.

Images was otherwise associated with, sponsoring, or endorsing Stability AI and its AI-generated images.[79] The Complaint also alleges trademark dilution, resulting from Stability AI's inclusion of a "Getty" watermark on AI-generated images that lack the quality of images that a customer would find on the Getty website.[80] Finally, the Complaint asserts that these improper uses cause both dilution by blurring (*i.e.*, lessening the capacity of Getty's mark to identify and distinguish goods and services) and by tarnishment (*i.e.*, by harming the reputation of Getty's mark by association with another mark).[81]

G. What About the GPTJudge and Their GPTLaw Clerk?

Finally, we are left to ask if it is permissible for judicial officers to use Chat-GPT or another GenAI system to research and/or draft opinions? At least three judges admit to having done so, asking the system "whether an autistic child's insurance should cover all the costs of his medical treatment,"[82] whether "an unusually high level of cruelty [in committing an assault and murder] should count against granting bail,"[83] and whether there was "any 'legitimate public interest' for journalists posting online photos of a 'woman showing parts of her body' without her consent."[84] At first blush, one might think, "what's the problem?" since we know that GPT-4, at least, passed the bar exam,[85] so "why not?"

The first concern is that ChatGPT can provide different

---

[79] See *A New Generation of Legal Issues: Part 2*, *supra* note 69.
[80] *Id.*
[81] *Id.*
[82] Luke Taylor, *Colombian judge says he used ChatGPT in ruling*, THE GUARDIAN (Feb. 3, 2023, 21:53 EST), https://www.theguardian.com/technology/2023/feb/03/colombia-judge-chatgpt-ruling. According to reports, ChatGPT concurred with the judge's final decision, responding "Yes, this is correct. According to the regulations in Colombia, minors diagnosed with autism are exempt from paying fees for their therapies." *Id.*
[83] Adam Smith, et. al., *Are AI chatbots in courts putting justice at risk?*, CONTEXT (May 4, 2022), https://www.context.news/ai/are-ai-chatbots-in-courts-putting-justice-at-risk.
[84] *Id.*
[85] Daniel M. Katz & Michael J. Bommarito, *GPT-4 Passes the Bar Exam*, SSRN (Mar. 15, 2023), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4389233.

answers to the same question at different times—if not hallucinate citations and other fictitious responses—and that it was trained on an unknown dataset from the Internet that contains no data past 2021.[86] But, there are other, more serious problems with this approach. If the judge or their clerk were to describe the facts and the law and prompt GenAI for the correct outcome, without independently verifying the accuracy of the information, this could raise an Article III judicial vesting-clause problem, since the U.S. Constitution Art. III §1 vests the judicial power of the United States in its federal courts and their duly appointed judges—not in AI. Even if the GenAI system were not being used to render the final decision in a case or controversy, and was instead used in a manner similar to how a judge or their clerk might undertake an Internet search concerning the facts in a case before them, this could easily run afoul of the American Bar Association's Model Code of Judicial Conduct Rule 2.9(C), which prohibits judges from independently investigating facts or considering facts not in the record or judicially noticed.[87] Using the GenAI system for independent research without informing counsel or providing them with an opportunity to object to arguments that are not in the record may very well expose the Court to sources of information that have not been put in evidence by the parties, or that raise other due process issues.[88]

Accordingly, the best advice we can give at this point is to exercise extreme caution—much like early advice concerning

---

[86] *See* OpenAI, *supra* note 27 ("Chat GPT is fine-tuned from a model in the GPT-3.5 series, which finished training in early 2022.").

[87] Model Rule 2.9(C) addresses Ex Parte Communications. It states that "A judge shall not investigate facts in a matter independently and shall consider only the evidence presented and any facts that may properly be noticed." ABA Model Code of Jud. Conduct: Canon 2.
https://www.americanbar.org/groups/professional_responsibility/publications/model_code_of_judicial_conduct/model_code_of_judicial_conduct_canon_2/rule2_9expartecommunications/.

[88] *See* ABA Standing Comm. on Ethics and Pro. Resp., Op. 478 (2017), https://www.abajournal.com/images/main_images/FO_478_FINAL_12_07_17.pdf. *See also* Avalon Zoppo, *ChatGPT Helped Write a Court Ruling in Colombia. Here's What Judges Say About Its Use in Decision Making*, NAT'L L. J. (Mar. 13, 2023),
https://www.law.com/nationallawjournal/2023/03/13/chatgpt-helped-write-a-court-ruling-in-colombia-heres-what-judges-say-about-its-use-in-decision-making/.

judicial use of social media—until a body of judicial ethics opinions is developed.

IV. WHAT THE FUTURE HOLDS

While we obviously have no crystal ball that can predict the future development of GenAI technology over the next few years, there is no doubt that it will revolutionize many fields, not the least of which will be the legal and justice systems. Generating fake but believable text, audio, and video of ordinary people spouting lies, misinformation, or defamatory content, committing crimes, or breaking the law will become feasible for just about any person with a working computer. So, too, will anybody be able to generate competent pleadings, in a matter of minutes, with great benefit to access to justice coming alongside the risk of many more vexatious filings flooding court dockets. As a result of these technological developments, our current approaches to managing cases and evidence may need to change. The legal status of AI-generated art (in particular, with respect to copyright eligibility, copyright infringement, and trademark infringement and/or dilution) will need to be resolved. Judges themselves will have to sort through AI-generated pleadings and arguments, including perhaps even using an AI clerk to filter out or respond to junk claims or imaginary citations (if and when this becomes possible). Judges may eventually join the revolution, using new GenAI systems to help them decide their cases or draft their opinions more effectively and efficiently, after problems involving inaccuracy and bias are resolved. And one day, judges may even be replaced by AI,[89] giving new meaning to the phrase "having one's day in court."

---

[89] Tara Vazdani, *From Estonian AI judges to robot mediators in Canada, U.K.*, THE LAWYER'S DAILY, https://www.lexisnexis.ca/en-ca/ihc/2019-06/from-estonian-ai-judges-to-robot-mediators-in-canada-uk.page (last visited Nov. 10, 2023). Indeed, OpenAI's release of the research and code for its new text-to-3D model, Shap-E—while we were in the midst of writing this piece—may even allow judges to be printed at some point! *See* Avran Piltch, *OpenAI's Shap-E Model Makes 3D Objects From Text or Images*, TOM'S HARDWARE (May. 4, 2023), https://www.tomshardware.com/news/openai-shap-e-creates-3d-models.