Synthetic Thought: Interrogating the Inscrutable

Will Lawrence*

Table of Contents

Introduction

The increasing use of artificial intelligence (AI) in the legal field raises too many complex legal and ethical issues to address in a single note. In *this* Note, I instead intend to focus in particular on AI's impact on legal practice and professional ethics with regard to the way such programs "reason" and the applicability of existing *Model Rules of Professional Conduct*.

Part I will first examine how (and whether) the adoption of AI affects core ethical obligations such as competence, diligence, confidentiality, and the duty to provide independent legal judgment.

* Georgetown University Law Center, J.D. expected 2026; The College of William & Mary, B.A. 2022;

Part II will then address the degree to which we have been anthropomorphizing and ascribing agency to generative algorithms, explaining the way they "think." Part II will next discuss the applicability of Model Rule of Professional Conduct 5.3, "Responsibilities Regarding Nonlawyer Assistance," in that context.

Finally, Part III will reflect on what lawyers ought to keep in mind when working with these tools, in hopes that the integration of generative AI into legal services serves to enhance the profession. In order to further facilitate that objective, this Note will argue that such technology should be adopted by the legal field (and that trying to avoid this outcome is a fool's errand), albeit subject to the same or similar rules to those the *Model Rules* already lay out.

As of February 2025, Formal Opinion 512 is the third-most recent formal opinion to have been issued by the American Bar Association (ABA).[1] It is the ABA's preliminary effort to address the use of generative artificial intelligence (GAI) tools, and as such, it is fairly broad ranging but necessarily shallow. The report notes that, "as with many new technologies, GAI tools are a moving target—indeed, a *rapidly* moving target—in the sense that their precise features and utility to law practice are quickly changing and will continue to change in ways that may be difficult or impossible to anticipate."[2] This is a mature stance to take on the matter, but perhaps a less than practical one. The law has historically been a slow-moving beast, and reactivity to evolving fields is therefore a persistently unfavorable matchup for it.[3]

I. SIGNPOSTING SMALLER SUBJECTS

A. COMPETENCE

Model Rule 1.1 reads, "A lawyer shall provide competent representation to a client. Competent representation requires the legal knowledge, skill, thoroughness and preparation reasonably necessary for the representation."[4] As for *whether* the use of GAI tools implicates this rule, one should hope the answer is obvious—of course it does. What, realistically, would *not* implicate it? Of far greater relevance is the question of *how*. Among Rule 1.1's requirements is that of technological competence:

> To maintain the requisite knowledge and skill, a lawyer should keep abreast of changes in the law and its practice, including the benefits and risks associated

1. ABA Comm. on Ethics and Prof'l Responsibility, Formal Op. 512 (2024). Neither Formal Opinion 513 nor 514 make direct reference to artificial intelligence, the former being concerned with "Duty to Inquire Into and Assess the Facts and Circumstances of Each Representation," ABA Comm. on Ethics and Prof'l Responsibility, Formal Op. 513 (2024), and the latter with "A Lawyer's Obligations When Advising an Organization About Conduct that May Create Legal Risks for the Organization's Constituents," ABA Comm. on Ethics and Prof'l Responsibility, Formal Op. 514 (2025).

2. ABA Comm. on Ethics and Prof'l Responsibility, Formal Op. 512, at 2 (2024).

3. Andrew M. Perlman, *The Legal Ethics of Generative AI*, SUFFOLK UNIV. L. REV. (forthcoming 2025).

4. MODEL RULES OF PROF'L CONDUCT R. 1.1 (2018) [HEREINAFTER MODEL RULES].

> with relevant technology, engage in continuing study and education and comply with all continuing legal education requirements to which the lawyer is subject.[5]

In the face of such rapidly evolving technology, is this a standard that the majority of lawyers are capable of meeting?

Where other comments on Model Rule 1.1 give caveats as to the requisite knowledge and skill required of lawyers, such as considering "the preparation and study the lawyer is able to give the matter,"[6] comment 8 seems to demand a great deal more from lawyers in the context of AI given the rapid pace at which the field continues to transform. This demand seemingly flies in the face of the rule's assertion that a lawyer "need not necessarily have special training or prior experience to handle legal problems of a type with which the lawyer is unfamiliar."[7]

Comment 8 was amended to specifically mention technology in 2012, when all the phrase "AI revolution" would have evoked in most were images of the *Terminator* franchise's Skynet.[8] Nevertheless, it was distinctly targeted at rapidly evolving and seemingly omnipresent technological advancements. In fact, it was added explicitly "in light of cloud computing and technology such as smartphones and tablets."[9] Over a decade later, it would not merely be flagrantly negligent to not have basic knowledge of such things in a professional context, it would be a severe handicap to one's existence in modern society altogether. In this context, it is tempting to dismiss concerns around how AI interfaces with the duty of competence as growing pains and nothing more. This, however, ignores an underlying component of that duty—understanding. According to a 2019 ABA resolution, under Model Rule 1.1, "lawyers also must have a basic understanding of how AI tools operate. While lawyers cannot be expected to know all the technical intricacies of AI systems, they are required to understand how AI technology produces results."[10] What this truly means, however, is less than clear. Notably, Model Rule 1.1 comment 8 did not *create* a duty of technology competence for lawyers; it merely made such a thing explicit.[11] As such, one can reasonably infer that AI does not demand anything new from lawyers as far as education is concerned. And yet, 'understanding how AI technology produces results' is an ambiguous requirement, the exact meaning of which is less than

---

5. MODEL RULES R. 1.1 cmt. 8.
6. MODEL RULES R. 1.1 cmt. 1.
7. MODEL RULES R. 1.1 cmt. 2.
8. THE TERMINATOR (Hemdale Film Corporation 1984).
9. *Applying today's legal ethics to today's AI (part 2)*, CASETEXT (Nov. 17, 2023), https://casetext.com/blog/ethical-use-ai-legal-2/ [https://perma.cc/3GUD-4W85].
10. ABA House of Delegates, Resolution 112, Aug. 13, 2019 [hereinafter ABA/HOD].
11. *Health Care Operations & Compliance, Overview - Duty of Technology Competence/Compliance Chart*, BLOOMBERG L., https://www.bloomberglaw.com/external/document/X2D4KAQ0000000/health-care-operations-compliance-overview-duty-of-technology-co [https://perma.cc/DPD9-FE86].

clear. This is one of the primary motivations in the legal field for the pursuit of explainability.

B. EXPLAINING EXPLAINABILITY

Sherlock Holmes would probably not be one of the most famous characters in all of Western canon if all his stories simply involved him arriving at the scene of a crime, giving it a once-over, and demanding a seemingly random witness be arrested. More to the point, even if his deduction might ultimately prove correct, it would be nothing short of irresponsible for Scotland Yard to comply with his demand based on this assessment alone. Nevertheless, this illustrates how artificial intelligence (AI) often works: capable of producing impressive results but not necessarily able to explain its thought process. Explainable Artificial Intelligence (XAI) is an effort to solve, or at least address, this 'black-box' conundrum.

While a precise definition is far from set in stone, XAI *generally* refers to a "set of processes and methods that allows human users to comprehend and trust the results and output created by machine learning algorithms."[12] At present, the most commonly called-for implementation would likely involve an interface as pictured above that can translate a model's often alien thought processes for a human audience. Ideally, this would make AI more accessible and accountable, especially in areas like healthcare, finance, and law, where understanding the reasoning behind a decision is essential.

Related to (but distinct from) explainability is the concept of "interpretability." Interpretability is the degree to which an observer can understand the cause of a *decision*. Explainability, meanwhile, looks at how the AI arrived at a given *result*.[13] Imagine a chef who has been cooking for years and can make an amazing dish just by tasting ingredients, with little concern for communicating the precise steps they took to anyone else. If one wanted to understand exactly why this chef added a certain spice or cooked something at a particular temperature, it could be difficult for them to explain their instinctive, step-by-step process. Interpretability is like asking this virtuoso chef to describe each ingredient they used and why they chose it. In the context of a neural network, this entails breaking down the network's decision-making process step-by-step to understand what each component part of the model is doing, like looking at individual ingredients to see how each one contributed to the final taste. Interpretability seeks to make the neural network's steps more transparent and easier to follow. If a neural network were used for approving loans, interpretability would help us understand why a specific applicant was approved, down to which factors made the most impact in the decision.

---

12. Violet Turri, *What is Explainable AI?*, CARNEGIE MELLON UNIV. SOFTWARE ENG'G INST. (Jan 17, 2022), https://insights.sei.cmu.edu/blog/what-is-explainable-ai/ [https://perma.cc/FE4N-ZE99].

13. *What is explainable AI?*, IBM, https://www.ibm.com/think/topics/explainable-ai [https://perma.cc/EUN6-2PMQ].

Meanwhile, explainability involves creating ways to clarify which ingredients (or features of the data) were most important in making the final dish (or the network's decision). It helps people get a general sense of how the network thinks and its key influences. For example, in image recognition, if a network identifies a dog in a picture, we want to know whether it recognized the dog because of the shape, the fur, or some other feature. One oft-cited story that illustrates the importance of this involves a program designed to differentiate dogs from wolves. When a husky dog was misidentified as a wolf, the developers probed the AI and determined that, because all the images of dogs the model had trained on had grassy backgrounds and all the images of wolves that the model trained on had snowy backgrounds, it was effectively ignoring the animals entirely and making its determination based on whether or not the background of the image was white.[14]

This example is fairly whimsical, but the same problem that underlies it also presents itself in circumstances of far greater consequence. An assessment of the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) risk assessment tool, which predicts the likelihood of a defendant reoffending, found that it was nearly twice as likely to falsely label Black defendants as high risk for reoffending compared to white defendants, while white defendants were more often incorrectly classified as low-risk.[15] This occurred despite the fact that, in order to "avoid decision bias based on race, gender, or appearance and to arrive at more objective outcomes," the system "does not use race, gender, employment, or living place" in its assessments.[16] In much the same way as a person is the product of all their experiences, an algorithm can only be as accurate as the data on which it is trained, and biased inputs will inevitably lead to biased predictions from the model. XAI may well prove vital in identifying these biases, in the model, in the data, and potentially even in our society.

It is important to remember that XAI is an inchoate effort, and finding measures that successfully accomplish the effort's goals has proven challenging.[17] It is so new, in fact, that even the basic terminology is in flux. Experts do not

---

14. *See Husky or Wolf? Using a Black Box Learning Model to Avoid Adoption Errors*, APPLIED INNOVATION (Aug 24, 2017), https://innovation.uci.edu/2017/08/husky-or-wolf-using-a-black-box-learning-model-to-avoid-adoption-errors/ [https://perma.cc/7EXN-77RA]; Henry S. Kenyon, *AI, Please Explain Yourself*, SIGNAL MEDIA (Feb 01, 2018), https://www.afcea.org/signal-media/ai-please-explain-yourself; Alexiei Dingli, *It's magic . . . I owe you no explanation!*, MEDIUM (Nov 20, 2018), https://medium.com/humain-technologies/its-magic-i-owe-you-no-explanation-51e4ba4a2337 [https://perma.cc/4KEF-UUD4]; Charles Simon, *AI Transparency will Lead to New Approaches*, INSIDEAI NEWS (Apr 21, 2020), https://insideainews.com/2020/04/21/ai-transparency-will-lead-to-new-approaches/ [https://perma.cc/7NLD-HNWS].

15. Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/TK4V-9BYH].

16. Hans de Bruijn, Martijn Warnier & Marijn Janssen, *The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making*, 39 GOV'T INFO. Q. 101666, at 3 (2022).

17. *See* Alessio Malizia & Fabio Paternò, *Why Is the Current XAI Not Meeting the Expectations?*, 66 COMMC'NS OF THE ACM no. 12 20, 20-21, 23 (2023).

necessarily agree with the above distinction between "explainability" and "interpretability," with some regarding the definitions as flipped and others using them interchangeably.[18] Nevertheless, XAI at least holds promise for creating fairer, more accountable AI systems. By offering insights into how models make decisions, XAI might well empower developers, policymakers, and society at large to address and correct biases, improving trust and ethical standards in AI. While current methods are still developing, ongoing research and collaboration are paving the way for more transparent and interpretable models that are both effective and just.[19] In this sense, XAI represents not only a tool for refining technology but also an opportunity to shape a more equitable and responsible future in AI applications.

C. CONFIDENTIALITY

Model Rule 1.6(a) states that a lawyer "shall not reveal information relating to the representation of a client unless the client gives informed consent, the disclosure is impliedly authorized in order to carry out the representation," or revealing said information is explicitly permitted in a later subpart of the rule.[20] Furthermore, the rule requires a lawyer to "make reasonable efforts to prevent the inadvertent or unauthorized disclosure of, or unauthorized access to, information relating to the representation of a client."[21] Of the current *Model Rules*, this may well be the most likely to seriously handicap lawyers' abilities to use GAI tools, or at least the ones currently commercially available.

In early 2023, technicians at Samsung's semiconductor division inadvertently leaked confidential information while using ChatGPT to check their source code.[22] By inputting the source code into the AI tool, the technicians effectively surrendered it as information with which to train OpenAI's language model, allowing it to regurgitate that information to any other user who asked for

---

18. *See, e.g.*, Sajid Ali et al., *Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence*, 99 INFO. FUSION 101805, at 2 (2023); Finale Doshi-Velez & Been Kim, *Towards a Rigorous Science of Interpretable Machine Learning* (Mar. 2, 2017) (unpublished manuscript), https://arxiv.org/pdf/1702.08608 [https://perma.cc/ES7S-K6XJ].

19. *See generally* Nitesh Upadhyaya, *Unlocking the Black Box: Advancements in Explainable AI and Model Interpretability*, 11 INT'L ADVANCED RSCH. J. IN SCI., ENG'G AND TECH. 160 (2024); Daniel Enemona Mathew et al., *Recent Emerging Techniques in Explainable Artificial Intelligence to Enhance the Interpretable and Understanding of AI Models for Human*, 57 NEURAL PROCESSING LETTERS 16 (2025).

20. MODEL RULES R. 1.6.

21. MODEL RULES R. 1.6.

22. *A Case Study on Samsung's ChatGPT Incident*, HUMAN FIREWALL, https://humanfirewall.io/case-study-on-samsungs-chatgpt-incident/[https://perma.cc/TJ5P-H6VE]; *see* Siladitya Ray, *Samsung Bans ChatGPT Among Employees After Sensitive Code Leak*, FORBES (May 2, 2023), https://www.forbes.com/sites/siladityaray/2023/05/02/samsung-bans-chatgpt-and-other-chatbots-for-employees-after-sensitive-code-leak/[https://perma.cc/4GXV-3UMJ]; Cecily Mauran, *Whoops, Samsung workers accidentally leaked trade secrets via ChatGPT*, MASHABLE (Apr. 6, 2023), https://mashable.com/article/samsung-chatgpt-leak-details [https://perma.cc/87QG-SP4B]; Vilius Petkauskas, *Lessons learned from ChatGPT's Samsung leak*, CYBERNEWS (May 9, 2023), https://cybernews.com/security/chatgpt-samsung-leak-explained-lessons/.

something like it. This trait is hardly unique to ChatGPT; large language models function based on learned data, so any GAI tool that works in a similar fashion presents the same risk.[23] Likewise, precisely the same risk of leaks exists in the legal field as in the technical one.

Were a lawyer to use a generative AI tool like ChatGPT to compose a draft of a contract or compose other legal documents, or have a tool read over, check, or revise a document they had written themselves, that input would become data accessible by said AI.[24] All of the information that the lawyer input, confidential or otherwise, would be annexed into ChatGPT's impossibly vast collection of data points.[25] Beyond this point, the toothpaste is out of the tube; data fed into the system cannot be 'recovered' any more than a grain of salt dissolved in the ocean. Furthermore, the outputted essay is also considered part of the collective.[26] If the lawyer had instead asked ChatGPT to modify their existing draft, the new version ChatGPT presented would be considered an output of the software and would therefore be treated as content that can be retained or otherwise transformed by the app.[27] Although OpenAI has since changed its terms of service such that its models will no longer use user data submitted through its API as the default option for training (unless a customer or organization opts in), this is reportedly inconsistent across versions.[28] And, again, ChatGPT is far from the only large language GAI tool on the market.

With these risks in mind, are lawyers entirely precluded from using such AI tools? Not literally speaking, no. However, there exists an extent to which lawyers' ability to use such tools might be hobbled such that using them at all would be rendered pointless. Without being able to apply facts of a given case or details of a given contract to a prompt, GAI tools would hardly be an improvement over templates, on which many in certain practice areas likely already rely, where

---

23. Petkauskas, *supra* note 22.

24. Lance Eliot, *Generative AI ChatGPT Can Disturbingly Gobble Up Your Private and Confidential Data, Forewarns AI Ethics and AI Law*, FORBES (Jan. 27, 2023, 8:00 AM), https://www.forbes.com/sites/lanceeliot/2023/01/27/generative-ai-chatgpt-can-disturbingly-gobble-up-your-private-and-confidential-data-forewarns-ai-ethics-and-ai-law/ [https://perma.cc/E2WH-MBJH]; *see* Mauran, *supra* note 22.

25. Eliot, *supra* note 24.

26. *Id.*

27. *Id.*

28. *See* Paul Hill, *OpenAI no longer uses API customer data to train its LLMs*, NEOWIN (May 5, 2023, 1:10 PM), https://www.neowin.net/news/openai-no-longer-uses-api-customer-data-to-train-its-llms [https://perma.cc/C3DP-DNUN]; *see also* bilalkareemseek, *How to use chatgpt without providing it with training data?*, COMMUNITY.OPENAI.COM (Mar 30, 2023, 8:58 PM), https://community.openai.com/t/how-to-use-chatgpt-without-providing-it-with-training-data/111514/3 [https://perma.cc/PC96-AKU3]; giancarloerra, *Option to not use your data to train models disappeared from Custom GPT options*, COMMUNITY.OPENAI.COM (Nov. 10, 2023, 5:43 AM), https://community.openai.com/t/option-to-not-use-your-data-to-train-models-disappeared-from-custom-gpt-options/492391 [https://perma.cc/HAA6-STRL]. *See generally* Franck Dernoncourt, *Does opting out of having my content used for improvement mean there are no other forms of data retention of my content by OpenAI?*, STACKEXCHANGE (Jan. 31, 2023, 8:54 AM), https://datascience.stackexchange.com/questions/118173/does-opting-out-of-having-my-content-used-for-improvement-mean-there-are-no-othe [https://perma.cc/94XU-86Y9].

relevant. Additionally, although Model Rule 1.15, which concerns the safekeeping of property,[29] does not explicitly apply to data (data not falling under the traditional legal scope of "property"[30]), carelessly handling sensitive information a client entrusts to their attorney is certainly at odds with the spirit of that rule.

II. SUPERVISION AND ANTHROPOMORPHISM

Finally, and potentially most concerning of all the ABA's initially-presented guidance, comes the issue of supervision. It was not so long ago that AI, in the most general sense, would refer to synthetic sapient minds capable of human-like thought, exemplified by science-fiction characters like *2001: A Space Odyssey*'s HAL 9000[31] or *System Shock*'s SHODAN[32]. AI, as it has been most commonly known in the past year, is not that.[33] AI, as it has been addressed in this note, is not that. The increasing reliance on generative artificial intelligence tools across every field, not merely that of law, has led to a widespread tendency to anthropomorphize these systems, attributing to them qualities such as 'thinking,' 'reasoning,' and even 'decision-making.' While these descriptors are often used metaphorically, their repeated use risks creating a perception of agency or independence that is fundamentally misleading. This section explores the nature of GAI's algorithmic processes, highlights the ethical implications of ascribing human-like qualities to these systems, and underscores the importance of maintaining clarity about their operational limitations.

Model Rule 5.3 concerns responsibilities regarding nonlawyer assistance and requires lawyers to oversee nonlawyers who help them provide legal services and to ensure their compliance with the other rules of professional conduct.[34] Among the amendments to the *Model Rules* in 2012 was one that changed the language of Model Rule 5.3 from "assistants" to "assistance."[35] This change was not one targeted at or anticipating AI; rather, it was merely meant to "highlight that lawyers have an obligation to make reasonable efforts to ensure that *all* nonlawyers that assist them act in a manner that is consistent with the attorney's professional obligations."[36] While broadening the scope of this rule is largely positive and likely seemed unproblematic at the

29. MODEL RULES R. 1.15.

30. *See* James Grimmelmann & Christina Mulligan, *Data Property*, 72 AM. U. L. REV. 829, 830–32 (2023).

31. 2001: A SPACE ODYSSEY (Stanley Kubrick Productions 1968). Consider HAL's own grandiose claims: "The 9000 series is the most reliable computer ever made. No 9000 computer has ever made a mistake or distorted information. We are all, by any practical definition of the words, foolproof and incapable of error," *id*.

32. SYSTEM SHOCK (LookingGlass Technologies 1994).

33. Fei-Fei Li & John Etchemendy, *No, Today's AI Isn't Sentient. Here's How We Know*, TIME (May 22, 2024, 2:36 PM), https://time.com/collection/time100-voices/6980134/ai-llm-not-sentient/ [https://perma.cc/X3TN-8QEE]; Cade Metz, *A.I. Is Not Sentient. Why Do People Say It Is?*, THE NEW YORK TIMES (Aug. 5, 2022), https://www.nytimes.com/2022/08/05/technology/ai-sentient-google.html.

34. MODEL RULES R. 5.3.

35. *See* ABA/HOD, *supra* note 10, at 6; ABA MODEL GUIDELINES FOR THE UTILIZATION OF PARALEGAL SERV.'S (AM. BAR ASS'N 2018), at n. 3, https://assets.osbplf.org/forms/practice_forms/ABA%20Model%20Guidelines%20for%20Utilization%20of%20Paralegal%20Services.pdf [https://perma.cc/4933-3NJ8].

36. *See* ABA MODEL GUIDELINES FOR THE UTILIZATION OF PARALEGAL SERV.'S, *supra* note 35.

time, this minor shift in language raises questions surrounding the use of GAI tools to which Model Rule 5.3 would not have been subject before the change. Regardless of the language we use to discuss GAI, few would go so far as to claim that such a tool rises to the level of an assistant—ChatGPT could hardly be paid for its work. Nevertheless, the change puts non-human legal assistance squarely within the ambit of the ABA's rules. This is not merely speculative; the ABA itself has said as much as early as 2019.[37] This leaves modern lawyers with a conundrum: To what extent must they "supervise" the tools assisting them in their work, above and beyond merely using them? Certainly, one could be said to receive "assistance" from any number of resources, but an encyclopedia need not be "supervised" because it does not act autonomously. Autonomous action, therefore, seems the most relevant thread at which to pull to determine whether Model Rule 5.3 is at all relevant and, if it is, how to proceed.

A. THE PROBLEMS WITH ANTHROPOMORPHIZATION, GENERALLY

AIs, as we know of them today, are mimics.[38] Even the Turing test, which has long been popularly considered the benchmark for machine intelligence, is an exercise in imitation.[39] Fundamentally, large language models function in the same underlying manner as the kind of predictive text with which many of us are much more familiar. Apple introduced its QuickType function with iOS 8 in 2014, before OpenAI had even opened its doors.[40] Despite this fact, the occasional person (often an engineer in the AI field itself) has confidently proclaimed that a truly conscious machine had been achieved, even before the 2023 incident in which Bing's chat engine tried to convince New York Times journalist Kevin Roose to leave his wife for it.[41] This is by no means a surprising outcome—humans have a natural tendency toward applying human qualities to nonhuman objects, even when mimicking humanity is not one of that object's intended functions.[42] But awareness of this phenomenon does little to countermand its effects on its own, and every field, not just law, would do well to avoid complacency. As Murray Shanahan, a senior scientist at Google DeepMind[43], puts it:

---

37. *See* ABA/HOD, *supra* note 10, at 6.

38. *See* Metz, *supra* note 33.

39. *See* Ben Garside, *How anthropomorphism hinders AI education*, RASPBERRY PI BLOG (Apr. 13, 2023), https://www.raspberrypi.org/blog/ai-education-anthropomorphism/ [https://perma.cc/5VN8-X5GM].

40. *Compare* APPLE, https://www.apple.com/my/ios/whats-new/quicktype/ [https://perma.cc/K2VD-GCM5], *with Introducing OpenAI*, OPENAI, https://openai.com/index/introducing-openai/ [https://perma.cc/J7ER-32PW].

41. Nir Eisikovits, *AI isn't close to becoming sentient – the real danger lies in how easily we're prone to anthropomorphize it*, THE CONVERSATION (Mar. 15, 2023, 8:22 AM), https://theconversation.com/ai-isnt-close-to-becoming-sentient-the-real-danger-lies-in-how-easily-were-prone-to-anthropomorphize-it-200525 [https://perma.cc/6YHZ-3CWZ]; *see* Metz, *supra* note 33. Kevin Roose, *A Conversation with Bing's Chatbot Left Me Deeply Unsettled*, THE NEW YORK TIMES (Feb. 16, 2023), https://www.nytimes.com/2023/02/16/technology/bing-chatbot-microsoft-chatgpt.html.

42. *See* Eisikovits, *supra* note 41.

43. Ian Sample, *'We can't compete': why universities are losing their best AI scientists*, THE GUARDIAN

> As we build systems whose capabilities more and more resemble those of humans, it becomes increasingly tempting to anthropomorphize those systems, even though they work in ways that are fundamentally different than the way humans work. . . . But it is a serious mistake to unreflectingly apply to AI systems the same intuitions that we deploy in our dealings with each other, especially when those systems are so profoundly different from humans in their underlying operation.[44]

B. HOW LLMS "THINK"

"LLM" here refers to "large language models," not Masters of Laws. This is another point that makes discussing AI particularly difficult for people in the legal field. Please, put the latter out of your mind for now. Large language models represent the bulk of popularly available GAI programs that generate text, including OpenAI's GPT, Meta's Llama, and Anthropic's Claude. A precise description of how these programs function is well beyond the *Model Rules*' requirements.[45] The powerful LLMs we have today are the culmination of multiple decades of AI research, and expecting a person outside of that field to understand the technical workings after only a short summary is far too much to reasonably ask of them. Nevertheless, the ABA seems to be under the impression that the requirement of technological competence is unbending in this regard,[46] which invites us to scrutinize why.

First, as the name suggests, LLMs deal exclusively with *language*. Other GAI models such as Stable Diffusion, which generates images, entered the public perception at much the same time as the current wave of popular LLMs, and there exists overlap in their functionality, but they are not necessarily the same. This note, concerned as it is with the legal field, will focus on LLMs in particular. GAI operates on a fundamental core of algorithms and data—for LLMs, a *staggering* amount of data.[47] Specifically, LLMs function using neural networks, which, in *very* basic terms, are highly advanced, complex forms of pattern recognition.[48] These algorithms do not generally interface with language *as* language. Rather, words and word fragments (collectively called "tokens") are converted into "vectors."[49] By

---

(Nov. 1, 2017, 06:30 AM), https://www.theguardian.com/science/2017/nov/01/cant-compete-universities-losing-best-ai-scientists [https://perma.cc/PL53-F6MX].

44. Murray Shanahan, *Talking About Large Language Models*, COMMUNICATIONS OF THE ACM, 67, 68 (2024).

45. *See generally* MODEL RULES R. 1.1 cmt. 2 (stating that a lawyer "need not necessarily have special training or prior experience to handle legal problems of a type with which the lawyer is unfamiliar," and therefore indicating that the ABA does not expect lawyers to become deeply learned on every topic they encounter).

46. *See* ABA/HOD, *supra* note 10, at 4.

47. John Edwards, *Can AI Ever Become Capable of Original Thought?*, INFORMATION WEEK (Oct. 30, 2023), https://www.informationweek.com/machine-learning-ai/can-ai-ever-become-capable-of-original-thought- [https://perma.cc/T7CD-CL66].

48. Ninad Kulkarni, *Building basic intuition for Large Language Models (LLMs)*, MEDIUM (Dec. 3, 2023), https://medium.com/@thefrankfire/building-basic-intuition-for-large-language-models-llms-91f7ca92dfe7 [https://perma.cc/MV6T-DWY6].

49. Timothy B. Lee & Sean Trott, *Large language models, explained with a minimum of math and jargon*, UNDERSTANDING AI (Jul. 27, 2023), https://www.understandingai.org/p/large-language-models-explained-with [https://perma.cc/6WXJ-V4A3].

analyzing the massive amount of data fed into it, the program determines 'spatial' relationships between tokens determined by how often tokens co-occur in proximity to one another. Language models represent vectors in hundreds or even thousands of dimensions, rather than the two or three to which the human mind is limited.[50] Comprehending how LLMs determine these relationships is a task for which our biological hardware is simply not equipped.

Despite this complexity of process, whether this could be considered actual "thought" is a matter of philosophical debate.[51] One thing is certain, however: this process cannot be described as comparable to conventionally human modes of thought. While AI can combine existing knowledge into new arrangements, it does not "think" in the same way humans do.[52] As Greg Kostello, CTO at Huma.AI, a healthcare AI company, explains, "It's more accurate to say that AI can exhibit 'emergent behaviors,' or produce results that weren't explicitly programmed."[53] Current AI systems are primarily based on pattern recognition and statistical inference from large datasets, but they cannot engage in truly original thinking.[54]

The possibility of machines acquiring true original thinking or some other form of consciousness is debated among experts in both the AI and neuroscience fields, and although predicting a timeframe is an intensely speculative business, "many experts believe we are several decades away, if not longer, from even the potential of such a development."[55]

This discussion is not intended to downplay AI's capabilities. LLMs are far from unsophisticated. Quite to the contrary, the recent developments in the AI field have been *intensely* impressive, and none of the foregoing should be used as an argument to the contrary. Nor should it be read as a luddite argument. It is merely intended to rebut the presumption that these algorithms are appreciably similar to conventional human thought—a presumption that, for many, may well be a quirk of language rather than sincere belief.

III. PROSPECTIVE SUPERVISORY DUTIES

With that in mind, how does one appropriately "supervise" something with thought processes so alien that some do not define them as thought processes at all?

A. HALLUCINATIONS

"Hallucination" is a general term for when an AI program generates flawed, false, or nonsensical information, often presenting it confidently as the truth.[56]

---

50. *Id.*
51. Edwards, *supra* note 47.
52. *Id.*
53. *Id.*
54. *Id.*
55. *Id.*
56. *See* Jonathan Ciottone, *A Lawyer's Guide to Understanding AI Hallucinations in a Closed System*, NAT'L L. REV. (June 26, 2024), https://natlawreview.com/article/lawyers-guide-understanding-ai-hallucinations-closed-

Note that, as with many definitions in the AI space, this definition is an evolving one, with no precise, concrete means of determining what a hallucination is and is not.[57] Nevertheless, by the time of writing this Note, stories of lawyers citing fictional cases in a legal brief have become almost commonplace.[58] Such errors have resulted in some lawyers facing disciplinary actions, including sanctions[59] and several-thousand-dollar fines.[60] As recently as February of 2025, attorneys from the firm Morgan & Morgan filed a document citing nine nonexistent cases in district court,[61] prompting the firm to send an "urgent" email to its more than 1,000 lawyers on the topic.[62] This issue is so front-of-mind in the legal field that even Chief Justice John Roberts felt the need to comment on it in his 2023 Year-End Report on the Federal Judiciary.[63] The pervasiveness of this worry is not unfounded; a 2023 study from Stanford's RegLab and the Institute for Human-Centered AI demonstrated that hallucination rates in response to legal queries ranged from 69% to 88% when asked to state-of-the-art language models.[64] And although the technology is accelerating quickly, the problem is deep enough that even tools tailored specifically toward legal research, including Lexis+ AI, Westlaw AI-Assisted Research, and Ask Practical Law AI, still hallucinate, albeit at a substantially lesser rate than general-purpose AI models such as GPT-4.[65]

The concerns with hallucination are not merely limited to GAI tools misrepresenting the truth. Because, in the broadest sense, LLMs become more powerful

---

system [https://perma.cc/2MAZ-WKP5].

57. *See* Lance Eliot, *AI Ethics Lucidly Questioning This Whole Hallucinating AI Popularized Trend That Has Got To Stop*, FORBES (Aug. 24, 2022, 8:00 AM), https://www.forbes.com/sites/lanceeliot/2022/08/24/ai-ethics-lucidly-questioning-this-whole-hallucinating-ai-popularized-trend-that-has-got-to-stop/ [https://perma.cc/X5XS-ZY5Z].

58. *See* Varun Magesh et al., *AI on Trial: Legal Models Hallucinate in 1 out of 6 (or More) Benchmarking Queries*, STANFORD UNIV. (May 23, 2024), https://hai.stanford.edu/news/ai-trial-legal-models-hallucinate-1-out-6-or-more-benchmarking-queries [https://perma.cc/N7CF-UEAF]; *see also* Benjamin Weiser & Jonah E. Bromwich, *Michael Cohen Used Artificial Intelligence in Feeding Lawyer Bogus Cases*, N.Y. TIMES (Dec. 29, 2023), https://www.nytimes.com/2023/12/29/nyregion/michael-cohen-ai-fake-cases.html.

59. Sara Merken, *Texas lawyer fined for AI use in latest sanction over fake citations*, REUTERS (Nov. 26, 2024, 8:20 PM), https://www.reuters.com/legal/government/texas-lawyer-fined-ai-use-latest-sanction-over-fake-citations-2024-11-26/; *see* Magesh et al, *supra* note 58.

60. *See* Sara Merken, *AI 'hallucinations' in court papers spell trouble for lawyers*, REUTERS (Feb. 18, 2025, 3:55 PM), https://www.reuters.com/technology/artificial-intelligence/ai-hallucinations-court-papers-spell-trouble-lawyers-2025-02-18/; Samantha Cole, *Lawyers Caught Citing AI-Hallucinated Cases Call It a 'Cautionary Tale'*, 404 MEDIA (Feb. 13, 2025, 12:33 PM), https://www.404media.co/lawyers-caught-citing-ai-hallucinated-cases-call-it-a-cautionary-tale/ [https://perma.cc/6X5U-A6AE].

61. Sara Merken, *Lawyers in Walmart lawsuit admit AI 'hallucinated' case citations*, REUTERS (Feb. 10, 2025, 6:35 PM), https://www.reuters.com/legal/legalindustry/lawyers-walmart-lawsuit-admit-ai-hallucinated-case-citations-2025-02-10/; *see* Merken, *supra* note 60.

62. Merken, *supra* note 60.

63. JOHN G. ROBERTS, JR., 2023 YEAR-END REPORT ON THE FEDERAL JUDICIARY 6 (2023).

64. Daniel E. Ho, Matthew Dahl, Varun Magesh & Mirac Suzgun, *Hallucinating Law: Legal Mistakes with Large Language Models are Pervasive*, STANFORD UNIVERSITY (Jan. 11, 2024), https://hai.stanford.edu/news/hallucinating-law-legal-mistakes-large-language-models-are-pervasive [https://perma.cc/2PMZ-6DUL].

65. *See* Magesh et al., *supra* note 58.

and more accurate when trained on more data, lawyers dealing with litigation in lower courts or in less prominent jurisdictions are liable to be particularly vulnerable to hallucinations.[66]

Taken together with the inhuman thought processes upon which GAI relies, the extent to which it can be "supervised" as a paralegal might be limited at best. At most, this form of supervision would be more akin to spinning a roulette wheel or rolling dice and having to fact-check the randomly generated result; whether this would represent an easing of a lawyer's workload is dubious and likely somewhat subjective.

B. IS SUPERVISION EVEN POSSIBLE? AN ALTERNATIVE FRAMEWORK

The problem with supervising extends beyond AI's inhumanity as well. As was brought up in the earlier explanation of XAI, the machine learning field has long faced a so-called "black box" problem, where the inner functions of an AI's decision-making process are obscure and hard to grasp.[67] If a program cannot be guided and influenced by the lawyer meant to be supervising it, then the supervision is necessarily going to suffer. This process is comparable to a lawyer allowing their AI 'assistant' to work unsupervised and vetting the results after the fact. And in the case of a black box AI, which systems like ChatGPT predominantly are, the lawyer would be incapable of finding out where or why errors even occurred.

This obfuscation points to the most glaring issue with the ABA's guidance—it asks that lawyers understand, to a certain extent, how these AI programs work, when that is at times a difficult task for even the experts.[68] The questions they propose that lawyers should be asking focus too greatly on the *what* and *how*, rather than the more instructive *when* and *why*. Rather than asking lawyers to learn about an LLM's functions, the ABA should instead suggest that lawyers be better informed as to LLMs' use-cases, and, perhaps more importantly, when their use should be avoided.

As such, there are certain practices it is incumbent upon any lawyer using these tools to remember while doing so. Bear in mind that these are strictly minimal practices—what should be required rather than what would simply be 'nice.' First, foremost, and perhaps obviously to some, is the necessity of checking the

66. *See* Ho et al., *supra* note 64.

67. *See* Lou Blouin, *AI's mysterious 'black box' problem, explained*, UNIV. MICHIGAN-DEARBORN (Mar. 6, 2023), https://umdearborn.edu/news/ais-mysterious-black-box-problem-explained [https://perma.cc/6WSS-FVZV]; Saurabh Bagchi, *What is a black box? A computer scientist explains what it means when the inner workings of AIs are hidden*, THE CONVERSATION (May 22, 2023), https://theconversation.com/what-is-a-black-box-a-computer-scientist-explains-what-it-means-when-the-inner-workings-of-ais-are-hidden-203888 [https://perma.cc/5L2F-WCJF]; *Cracking the Code: The Black Box Problem of AI*, SCADS AI DRESDEN/LEIPZIG (Jul. 19, 2023), https://scads.ai/cracking-the-code-the-black-box-problem-of-ai/ [https://perma.cc/Y3DV-F27N]; Matthew Kosinski, *What is black box AI?*, IBM, https://www.ibm.com/think/topics/black-box-ai [https://perma.cc/4HYS-M53Z].

68. *See* Blouin, *supra* note 67; Bagchi, *supra* note 67; Kosinski, *supra* note 67.

AI's outputted 'work.' This may, at first, seem counterintuitive; surely using the AI should require *less* of a time and effort investment, not more. However, one must consider the alternative of manually searching for a given case rather than offloading the task to an AI. Despite their hallucinatory faults, large language models have incredible potential as plain-language search engines if properly tuned to the task.[69] Should a case the AI retrieves actually exist, the time one would have taken to find it will likely have been drastically reduced compared to searching by more conventional means; it is meaningfully worth reinvesting that time to ensure the source's veracity. Second, one should remember to consider alternatives. If fact-checking an AI's output seems like too much of a pain, prior means of electronic search remain as available as they ever were. Even when wielding an AI hammer, one must recall that not every problem is a nail. Making a decision between these two approaches is a largely personal process, particular to both a given situation and the given attorney facing it.[70] However, this balancing act is of paramount importance for any lawyer who wishes to remain reliable. Perhaps bidding farewell to the Boolean search will be sufficient motivation for one lawyer to use AI tools, with all the new requirements for safety and accuracy that entails, while another might find reviewing their chatbot's homework more irritating than whatever time they might save is worth.

CONCLUSION

The integration of generative artificial intelligence (AI) into legal practice represents both an extraordinary opportunity and a profound ethical challenge. As the legal profession increasingly turns to AI tools to streamline processes, analyze data, and even assist in drafting legal documents, it is imperative to address the ethical implications head-on. This note has examined the ways in which generative AI impacts core professional obligations, including competence, diligence, and independent legal judgment, while also exploring the risks associated with anthropomorphizing these systems and ascribing them undue agency. Generative AI is not just a tool; it is a reflection of how innovation intersects with professional responsibility. As legal professionals, we must approach its integration thoughtfully, ensuring that our fundamental obligations to clients, courts, and society remain steadfast in an age of rapid technological change. Interfacing with these technologies requires careful analysis and constant vigilance—the kind that might have alerted a reader to the fact that this entire paragraph (save for this sentence) was itself written by ChatGPT.

---

69. *See* David Ellis, *New AI-based natural language feature makes complex searches in Dimensions faster and easier*, DIGIT. SCI. (Nov. 7, 2024), https://www.digital-science.com/news/new-ai-based-natural-language-feature-in-dimensions/ [https://perma.cc/3LNB-HW4J].

70. *See also* Vishal Lingineni, *Beyond the Search Box: Large Language Models (LLM's) vs. Search Engines and Unlocking LLM's Potential*, MEDIUM (Nov. 7, 2023), https://vishal-lingineni.medium.com/beyond-the-search-box-large-language-models-llms-vs-f709224c9d1b [https://perma.cc/ZV8V-B8LU].

It would be wonderful to end this note with a firm, robust conclusion and chart a clear course of action for the years to come. It would also be irresponsible. As is the case with many of the topics discussed thus far, the field is evolving at such a rapid pace that elements of this note may have a very short shelf life. Although some speculate that AI progress has been slowing down of late, that too is a difficult matter to call with any certainty.[71] But this note's potentially quick expiration is not to its detriment—the legal field has historically moved at a crawl, often loath to make small, immediate adjustments to compensate for the way the world changes. This approach has its merits, but the pace at which the world is changing is increasing, and quickly. If the law does not attempt to adjust itself to that pace, it may end up snapping under the resulting tension.

71. *See* Megan Sauer, *Google CEO: AI development is finally slowing down—'the low-hanging fruit is gone'*, CNBC (Dec. 8, 2024), https://www.cnbc.com/2024/12/08/google-ceo-sundar-pichai-ai-development-is-finally-slowing-down.html [https://perma.cc/2MHN-R39L].