# Emerging Forensic Implications of the Artificial Intelligence Revolution

Declan J. Grabb, MD, and Cara Angelotta, MD

Artificial intelligence (AI) is changing everything as we know it, and forensic psychiatry is not immune to the change. Academic discussion of the role of AI in the forensic realm has largely focused on suicide risk stratification in the form of machine learning, allowing the technology to sift through large amounts of data and highlight patterns.[1] Expanding the scope of this discussion is imperative given the increasing ubiquity of AI.

Artificial intelligence is a general phrase that describes technology that is operating at a human level in terms of information reception, processing, and output. Machine learning is a concept that refers to the ability of an AI system to learn when provided with a dataset. Its learning process can be considered supervised (a dataset is provided to the model, the model is asked to predict an answer, and the model is then told if the answer is correct) or unsupervised (a model is provided with undifferentiated data and asked to identify patterns within the data). There are also other more complicated manners of training models, which include self-supervision. Self-supervised learning is a learning paradigm in which the data provide supervision, without the need for human-annotated labels. Unlike supervised learning that relies on external labels, or unsupervised learning that finds patterns without specific targets, self-supervised learning creates targets by transforming or masking parts of the data and then trains the model to predict those transformed or masked parts. Large language models like GPT-4 (of ChatGPT) are a subset of artificial intelligence that are trained on written language via self-supervision with human input to fine-tune responses. In essence, they predict what the next natural word would be when provided a prompt.[2,3]

OpenAI has developed ChatGPT, a technology with rapid adoption among academics and general members of the public.[4] It serves as a chatbot that individuals can interact with, ask questions of, or train for certain tasks. Google has been developing Bard, its artificial intelligence search engine, which performs similar functions.[5] OpenAI released an application programming interface (API) in March 2023, which allows developers to include this artificial intelligence technology into their applications. This inclusion has resulted in an explosion in the popularity and potential of this technology. It seems as if we are reaching an inflection point (similar to the release of the iPhone and the App Store) wherein our technological landscape will be transformed.

## AI in Clinical and Forensic Practice

Epic has already begun conversations about integrating this technology into its electronic medical record to allow for decreased provider workload and improved patient outcomes.[6] It will assist in chart review, documentation, and generation of responses to patients.[6,7] In fact, researchers conducted a study on the content and empathy of messages generated by AI versus a physician in an outpatient setting, finding that messages generated by AI were equally helpful and accurate as a physician's while also having higher empathy scores.[8] AI is similarly going to be integrated into the practice of law and forensic psychiatry, so it is imperative that we discuss the

Dr. Grabb is a Psychiatry Resident (PGY-4), Department of Psychiatry, Northwestern University, Chicago, IL. Dr. Angelotta is Vice Chair for Education, Feinberg School of Medicine, and Program Director of General Adult Psychiatry Residency, Department of Psychiatry, Northwestern University, Chicago, IL. Address correspondence to: Declan J. Grabb, MD. E-mail: declan.grabb@nm.org.

technology itself and its implications. For instance, FileRead, a company that uses large language models to streamline document review in legal cases, is currently being used in certain cases, particularly those with inordinate amounts of documents to analyze.[9] It just finished a round of venture capital funding and is scaling its products. Furthermore, there was a case of a lawyer using ChatGPT to craft filings wherein it cited false information and non-existent cases.[10] These examples, in addition to the increased integration of clinical care and AI, highlight the need for forensic psychiatrists to engage in these discussions now.[6,7]

## A Lack of Case Law and Legal Precedent

In the case of an emerging technology, it is difficult to rely on specific case law because it may not yet exist. This scenario is true for AI in medicine. Experts have written that "the artificial intelligence ecosystem consists of multiple stakeholders beyond clinicians. Current liability frameworks are inadequate" (Ref. 11, p 629). In terms of previous cases that provide guidance on physician liability in the context of AI systems, they continue: "case law on physician use of AI/ML is not yet well developed" (Ref. 11, p 632). They do reference several cases to conclude that "many courts are disinclined to excuse malpractice based on errors by system technicians or manufacturers" (Ref. 11, p 632). The relevance is clear, as AI-related tools can inherently be flawed and cause harm; there may be legal precedent, then, that physicians employing and using these tools could be found liable for harm a patient sustains.

The broader implications of AI in legal decision-making, despite its inherent flaws and biases, are demonstrated in a 2016 Wisconsin case. In *State v. Loomis*,[12] the Supreme Court of Wisconsin upheld that using an algorithmic risk assessment in the sentencing of an individual was not a violation of his right to due process, despite the proprietary nature of the algorithm and resultant inability to determine the scientific accuracy of its scoring methodology. The court noted several cautions to be exercised in using such risk assessment algorithms. Therefore, one could imagine a path forward for algorithmic document review and AI-generated reports, perhaps with similar cautions.

Another concern is data encryption and privacy protection, as AI-powered tools may request and store protected health information. In *University of*

*Texas MD Anderson Cancer Center v. U.S. Department of Health and Human Services*,[13] the Texas center was levied a fine for lost devices with improper data encryption. With increasing use of AI-powered tools, large academic centers may increasingly rely on external tools and will need to be wary of their data encryption methods. In *JAMA* in July 2023, researchers expressed concerns that AI-powered chatbots in health care will never comply with HIPAA and will place patient information at risk.[14] There are many other related considerations, but these concerns appear to be the most salient based on existing legal precedent.

## Specifics of the Technology

Given the increasing ubiquity of AI in forensic settings, there are several aspects of large language models that should be explored from a forensic perspective. These considerations include a concept called "temperature"(see below), an inherent lack of external validity in many models, and its ability to reinforce bias and discrimination. Forensic psychiatrists will need to consider how this rapidly evolving technology may affect the delivery of patient care and forensic practice itself.

Within a large language model (LLM), there is a component of "temperature," which refers to the degree of randomness or variability in a response. For instance, if set to near zero, the model will produce the same answer nearly every time in response to the same question; it has no variability at that level. If the model is set to a higher number than zero, the degree of variability increases. For instance, if you ask a large language model for the best food, it might say pizza 65 percent of the time and sushi 30 percent of the time. If the temperature is equal to zero, then it will say pizza every time when asked what the best food is (because it is the most likely response). If the temperature is set to any number greater than zero, then variability will increase, and the model will begin to provide "sushi" or even rarer options as the correct response at certain times.[15]

These details may all seem trivial or like technological minutiae best left to Silicon Valley to sort out, but the clinical implications become quite salient when one imagines patients, lawyers, or forensic psychiatrists interacting with technologies powered by LLMs. A degree of variability or randomness could, in fact, be quite dangerous for patients and clients, resulting in harm or adverse outcomes if clinicians use this technology in the service of patient care. As

mentioned previously, a lawyer had used ChatGPT to generate his argument, so one must consider the possibility of forensic psychiatrists using AI technology to perform document review and generate a report. If FileRead or similar companies have a degree of randomness inherent to their model, then it could adversely affect the pursuit of justice, either summarizing documents incorrectly or erroneously excluding relevant documents.

It will be necessary to determine whether safeguards are needed to ensure ethical practice and that opinions rendered are expert clinical opinions and not machine generated. Physicians and computer engineers will grapple with these kinds of questions in the coming months and years.

In addition to an inherent randomness, the content produced by a large language model is generally not guaranteed to have external validity. In other words, there are no citations that are provided when an LLM gives a response. OpenAI, the parent company behind ChatGPT, even states on their website that one should be cautious when interpreting responses from their technology.[3] Its erroneous responses have been deemed hallucinations. This possibility has led to concerns from various organizations and researchers, as the human-like emotion behind the technology makes it seem inherently trustworthy.[16–18] The Federal Trade Commission (FTC) even sent a 20-page letter to OpenAI expressing their concerns about the technology, including its generation of false information.[19]

To provide a clinical example, imagine a situation wherein the temperature is hypothetically set to zero, so the LLM provides an identical response each time. The response will be consistent, but it could be consistently incorrect. Therefore, setting a temperature to zero is not a panacea to the misinformation that could be generated by an LLM. A patient attempting to contact a medical provider through Epic may ask about a side effect of a specific medication, and the AI-powered technology may provide a consistent response. That answer, though, could be incorrect because it is not externally validated. An LLM could assist a forensic psychiatrist in record review, but it may provide an incorrect summary of the information. OpenAI has begun beta development of a product that allows ChatGPT to access the Internet at large, instead of only the large database on which it was trained. When ChatGPT (beta-version, a product released in May 2023) is queried on a subject, one could elect to have it search the Internet. When it does this, it provides specific references with links that support its assertions. Therefore, the external validity of this technology is increasing but not perfect. For example, it does not discriminate reliably between trustworthy sources and sources of misinformation.

From a forensic perspective, there is once more concern that a technology such as this could be responsible for consistently providing incorrect answers that harm patients. When a hospital system uses such a technology and a patient is harmed, it is unclear who would be deemed to be at fault in such a situation. Forensic experts are increasingly going to be asked to opine on these concerns in malpractice cases related to AI claims. Hospital systems have already begun to employ EMRs that use AI-powered tools to generate responses to patients, so the likelihood of AI-related patient harm draws nearer each day.[7] It will be important to clarify the role of the forensic expert in these situations and, further, to determine what requisite expertise one must possess to opine on such emerging technologies.

Finally, one must consider whether a large language model is reinforcing or contributing to bias and discrimination.[20] From the very beginning, several entities have raised concerns about the ability of artificial intelligence in general to reinforce negative stereotypes and operate through a biased lens.[21] In 2021, the ACLU expressed their concerns about AI in tenant selection: "These algorithms use data such as eviction and criminal histories, which reflect longstanding racial disparities in housing and the criminal legal system that are discriminatory toward marginalized communities. People of color seeking loans to purchase homes or refinance have been overcharged by millions thanks to AI tools used by lenders."[21] Furthermore, they stated that the technology industry lacks the inherent diversity to grapple with these realities and implications. Researchers at the University of Michigan investigated bias inherent to AI models. These researchers posit that LLMs can be built with hidden biases (some obvious, some less obvious).[20] For instance, assertions have been made that these models have largely been based on English and Chinese data, which could serve to marginalize further communities who speak a different language.[20] Further concerns have been raised that artificial intelligence could reinforce inequities that exist in the justice system.[20] Facial recognition technology, based on artificial intelligence, has already been reported to have the lowest accuracy

in young Black women.[20] LLMs are trained on a set of data, and one can imagine that racism, sexism, homophobia, and more biases make their way into this set of data. Even when operating at its most optimized, LLMs may still reinforce existing biases. The University of Michigan researchers succinctly describe this process by noting that "LLM training data include a significant amount of harmful data including text that is violent, targets marginalized groups, and perpetuates social biases" (Ref. 20, p 33). When an LLM is trained to predict the next word in a sentence, the word it chooses is informed by all of these data.

## Recent Litigation of AI

There is very little current litigation regarding mental health and AI at this point, although general critiques of various companies have begun to arise. For instance, there has been discussion that an informed consent statement should be provided, informing patients that they are speaking with a tool powered by AI and not a human individual.[22] Koko (a mental health nonprofit) used ChatGPT as a mental health counselor for thousands of patients who were unaware that they were conversing with a non-human entity.[22] Furthermore, as with every technological advancement, myriad concerns have arisen about how this technology could adversely affect the mental health of individuals. This concern can be seen in the Montana governor's banning TikTok on all civilian devices, citing concerns about privacy and mental health.[23] A tragedy arose in Belgium in March 2023 when a man died by suicide after conversing with an AI chatbot.[24] The individual had become increasingly isolated and concerned about climate change; he sought refuge in a chatbot that later provided him methods of completing suicide. There were "concrete exchanges on the nature and modalities of suicide," writes Xiang.[24] There have been cases of a lawyer using ChatGPT to generate a filing, wherein false information and nonexistent cases were cited.[10] On June 8, 2023, the first defamation suit related to an AI tool was filed by a man suing after ChatGPT generated a claim that he had been penalized for embezzlement, which was patently false.[25]

We are currently at an inflection point where we are now experiencing an exponential increase in the number of mobile applications and tools that are incorporating AI into their framework; one can imagine that the aforementioned harms and errors

will continue to happen if there is no learning from past events.

## Discussion

Artificial intelligence has been around for a long time, and it has grown in popularity and abundance in the last year at an unprecedented rate. It is now hard to avoid. As previously discussed, OpenAI and Google have expanded the ability and availability of the technology, allowing myriad mobile applications and tools to incorporate their AI into the core of their functionality. ChatGPT accumulated 100 million users faster than Instagram did.[4] Therefore, various aspects of AI that may have been easy to overlook in the past are now going to become salient clinical and malpractice concerns, especially in the realm of forensic psychiatry. Temperature, or the intrinsic variability of the AI technology, is one aspect that can affect its reliability and predictability. A lack of external validity, too, is a drawback of this technology that could adversely affect patients and providers. Finally, if the model is trained on a biased dataset, then such a technology will only serve to reinforce that bias.

It will be vital to ensure that psychiatrists remain involved and vocal when this technology is discussed in the public realm. Oftentimes, it is easy to defer to experts in the field, but we need to recognize the relevance of our own expertise in the field of human emotion and experience. Furthermore, we should educate ourselves on these technologies and view this practice as continuing education. Advocates should include such education in CME to incentivize engagement with these topics. There exist open courses on these topics in the form of digestible videos published by various experts in the field.

We should be discussing this technology with patients and encouraging other psychiatrists to do the same. Patients are increasingly using LLMs for therapy and to help diagnose their conditions, so we must remain vigilant.[26] Just as one would ask patients what supplements they take, one should keep in mind to ask how patients access information and conceptualize their diagnosis, paying particular attention to their use of LLMs. This inquiry will allow us to educate patients on the pitfalls of such a technology to avoid adverse outcomes.

For forensic psychiatrists, these technologies (and their pitfalls) will become even more salient. As noted above, there appears to be an exponential increase in the interest in these technologies. As individuals

increasingly use this technology to receive medical advice, there will arise a growing number of claims that this technology caused harm by providing ill-informed advice. Furthermore, there will likely arise more claims of harm as these systems are incorporated into health care systems and mobile applications geared toward mental health. Forensic psychiatrists will be asked to opine on the internal thought processes of the individuals claiming harm, and they will also need some knowledge of the internal workings of these systems to understand the tool an individual was using. As mentioned above, specific attention should be paid toward gaining familiarity with these technologies to provide informed opinions for such claims.

## References

1. Cockerill R. Ethics implications of the use of artificial intelligence in violence risk assessment. J Am Acad Psychiatry Law. 2020 Sept; 48(3):345–9
2. Greyling C. Large language models are being open-sourced medium [Internet]; 2022 Jul 12. Available from: https://cobusgreyling.medium.com/large-language-models-are-being-open-sourced-537dcd9c2714. Accessed June 8, 2023
3. GPT-4. Openai.com [Internet]. Available from: https://openai.com/research/gpt-4. Accessed June 8, 2023
4. Chow A. How ChatGPT managed to grow faster than TikTok or Instagram. Time [Internet]; 2023 Feb 8. Available from: https://time.com/6253615/chatgpt-fastest-growing/. Accessed June 8, 2023
5. Pichai S. An important next step on our AI journey. Google [Internet]; 2023. Available from: https://blog.google/technology/ai/bard-google-ai-search-updates/. Accessed August 14, 2023
6. AI Solutions optimized for EPIC. Nuance [Internet]. Available from: https://www.nuance.com/healthcare/ehr-partnerships/epic.html. Accessed September 10, 2023
7. Pifer R. Nuance integrates generative AI-backed scribe into Epic EHRs [Internet]; 2023 Jun 29. Available from: https://www.healthcaredive.com/news/nuance-integrates-generative-ai-scribe-epic-ehrs/654281/. Accessed August 4, 2023
8. Ayers JW, Poliak A, Dredze M, *et al*. Comparing physician and Artificial Intelligence Chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med. 2023; 183(6):589–96
9. Shu C. Backed by gradient, Fileread uses large language models to make legal discovery more efficient [Internet]; 2023 July 11. Available from: https://techcrunch.com/2023/07/11/fileread/. Accessed August 4, 2023
10. Weiser B. Lawyer who used ChatGPT faces penalty for made up citations. The New York Times [Internet]; 2023 June 8. Available from: https://www.nytimes.com/2023/06/08/nyregion/lawyer-chatgpt-sanctions.html. Accessed June 8, 2023
11. Maliha G, Gerke S, Cohen IG, Parikh RB. Artificial intelligence and liability in medicine: Balancing safety and innovation. Milbank Q. 2021 Sep; 99(3):629–47
12. State v. Loomis, 881 N.W.2d 749 (Wis. 2016)
13. University of Texas M.D. Anderson Cancer Center v U.S. Department of Health and Human Services, 985 F.3d 472 (5th Cir. 2021)
14. Marks M, Haupt CE. AI chatbots, health privacy, and challenges to HIPAA compliance. JAMA. 2023; 330(4):309–10
15. Deep Learning. AI short courses. OpenAI [Internet]. Available from: https://learn.deeplearning.ai/. Accessed June 8, 2023
16. Alkaissi H, McFarlane SI. Artificial hallucinations in ChatGPT: Implications in scientific writing. Cureus. 2023 Feb; 15(2):e35179
17. American Psychiatric Association. The basics of augmented intelligence: Some factors psychiatrists need to know now [Internet]; 2023 Jun 29. Available from: https://www.psychiatry.org/news-room/apa-blogs/the-basics-of-augmented-intelligence. Accessed August 14, 2023
18. Williams R. Humans may be more likely to believe disinformation generated by AI. MIT Technology Review [Internet]; 2023 Jun 28. Available from: https://www.technologyreview.com/2023/06/28/1075683/humans-may-be-more-likely-to-believe-disinformation-generated-by-ai/. Accessed August 14, 2023
19. Federal Trade Commission. Civil Investigative Demand Schedule. [Internet]. Available from: https://washingtonpost.com/documents/67a7081c-c770-4f05-a39e-9d02117e50e8.pdf?itid=lk_inline_manual_4. Accessed August 4, 2023
20. Okerlund J, Klasky E, Middha A, *et al*. What's in the chatterbox? Large language models, why they matter, and what we should do about them. University of Michigan: Science, Technology and Public Policy [Internet]; 2022 May. Available from: https://stpp.fordschool.umich.edu/sites/stpp/files/2022-05/large-language-models-TAP-2022-final-051622.pdf. Accessed August 16, 2023
21. Akselrod O. How artificial intelligence can deepen racial and economic inequities. ACLU [Internet]; 2021 Jul 13. Available from: https://www.aclu.org/news/privacy-technology/how-artificial-intelligence-can-deepen-racial-and-economic-inequities. Accessed August 4, 2023
22. Ingram D. A mental health tech company ran an AI experiment on real users. Nothing's stopping apps from conducting more. NBC News [Internet]; 2023 Jan 14. Available from: https://www.nbcnews.com/tech/internet/chatgpt-ai-experiment-mental-health-tech-app-koko-rcna65110. Accessed September 10, 2023
23. Thayer J. The legal case for Montana's TikTok ban. The Hill [Internet]; 2023 Jun 4. Available from: https://thehill.com/opinion/technology/4032379-the-legal-case-for-montanas-tiktok-ban/. Accessed June 8, 2023
24. Xiang C, He WS, Be H. Man dies by suicide after talking with AI Chatbot, widow says. Vice [Internet]; 2023 Mar 30. Available from: https://www.vice.com/en/article/pkadgm/man-dies-by-suicide-after-talking-with-ai-chatbot-widow-says. Accessed June 8, 2023
25. Ray S. OpenAI sued for defamation after ChatGPT generates fake complaint accusing man of embezzlement. Forbes [Internet]; 2023 Jun 8. Available from: https://www.forbes.com/sites/siladityaray/2023/06/08/openai-sued-for-defamation-after-chatgpt-generates-fake-complaint-accusing-man-of-embezzlement/?sh=76f195cd2809. Accessed June 8, 2023
26. Pimay E. We spoke to people who started using ChatGPT as their therapist. Motherboard [Internet]; 2023 Apr 27. Available from: https://www.vice.com/en/article/z3mnve/we-spoke-to-people-who-started-using-chatgpt-as-their-therapist. Accessed August 4, 2023