ANALYSIS

Does AI Infringe IP Laws When It Uses Copyrighted Material for 'Inspiration'?

AI chatbots that have come into wide use in the past months, like ChatGPT, have caused a massive stir in their own right, demonstrating that sentience is not required for AI technology to be massively disruptive and to raise a host of complex social, business and legal questions.

March 27, 2023 at 09:25 AM

Special Sections

By Joseph Casino and Michael J. Kasdan | March 27, 2023 at 09:25 AM

Last year, an artificial intelligence engineer at Google caused a massive stir when he claimed that an AI chatbot he was experimenting with was “sentient.” (See https://www.scientificamerican.com/article/google-engineer-claims-ai-chatbot-is-sentient-why-that-matters/). While AI tech has not yet brought us sentient artificial intelligence, i.e., AI that is not just following a programmed algorithm, AI chatbots that have come into wide use in the past months, like ChatGPT, have caused a massive stir in their own right, demonstrating that sentience is not required for AI technology to be massively disruptive and to raise a host of complex social, business and legal questions.

The current AI implementations follow steps and algorithms programmed by humans. (See //www.uspto.gov/sites/default/files/documents/USPTO_AI-Report_2020-10-05.pdf at ii) (In a USPTO survey, the “majority viewed the concept of artificial general intelligence (AGI)—intelligence akin to that possessed by humankind and beyond—as merely a theoretical possibility that could arise in a distant future.”) When the AI is tasked to create something, whether that is a piece of art, an essay, a legal brief, a song, a screenplay, or a piece of software code, it needs to look at other material for a starting point, i.e., “inspiration” for what it is creating, so to speak. This raises interesting legal issues when such source material, often called training materials, is copyrighted or trademarked. If the source material is not copied exactly, can the AI be considered a copier or can it escape infringement because what it creates is a transformative independent development? Following very closely in time to the release of virally popular generative AI apps, we are now seeing a first wave of lawsuits that have been filed to test these very questions.

In the art space, Stability AI provides an AI-image generator called DreamStudio, which relies on a database of billions of images to create new images requested by users. Getty Images filed a lawsuit in the High Court of Justice in London against Stability AI claiming that training its AI with Getty’s copyrighted images (i.e., “data scraping”) would require a license from Getty Images. In the US, a group of artists sued Stability AI in the Northern District of California alleging copyright infringement by the generative app-maker. (See https://www.reuters.com/legal/transactional/lawsuits-accuse-

ai-content-creators-misusing-copyrighted-work-2023-01-17/) (For an overview of the complaint in this action, see: https://twitter.com/michaelkasdan/status/1614397752733061120). Most recently, Getty filed a copyright and trademark lawsuit in the District of Delaware, also against Stability AI.

Similar theories are the basis of lawsuits in the software development field, where Microsoft's Github, Inc. and OpenAI were sued in a class action lawsuit alleging copyright infringement because their GitHub Copilot and OpenAI Codexe use AI to create new programs based on a large database of publicly available, but copyrighted software.

Microsoft has moved to dismiss the case arguing that Codex merely generates coding suggestions in response to a person's requests. "A programmer can provide a short text request (e.g., "create a button on a website that lets a user upload a document"), and Codex will generate a coding suggestion to meet the request." Microsoft argues that plaintiff's allegations that the code they generate is similar to the copyrighted code are insufficient because no specific examples of copying have been provided in the complaint.

These lawsuits raise a novel legal issue: how should copyright law consider the training of AI systems using material that is copyrighted. In considering this, it is important to note that there is a lack of detailed public information of how applications like DreamStudio, Copilot and Codex generate their work to provide the AI generated results. The Andersen and Getty U.S. copyright lawsuits against Stability AI both contain their allegations as to how Stability AI functions. But these are allegations. Some important issues include, for example, whether in creating artwork does DreamStudio ever copy segments of prior works into a new work either literally or with some minor deviations. Without more information on the so-called "stable diffusion" algorithm that is used to create the artwork, such copying may not to qualify as independent creation particularly in cases where "similarities between [the] works are sufficiently overwhelming and pervasive (including perhaps unusual common errors) so that the similarities in themselves necessarily preclude the possibility of independent creation and thereby render the defendant's denial of copying a nongenuine issue of fact. See 3 Nimmer on Copyright Section 12.10 (2022). In order to show independent creation, the providers of the AI may need to provide clear explanations of how its AI created the alleged infringing work, a task that may be difficult for even the provider to trace and explain if its AI adds to or modifies its basic algorithm as it receives more training data.

While these legal issues are novel for copyright, consideration has also been given to ensuring that there can be a tracing and explanation of what AI did to provide its output. For example, the European Union has provided ethical guidelines on AI to require transparency, requiring that algorithms be traceable, explainable and free from bias. (See, e.g., EU guidelines on ethics in artificial intelligence: Context and implementation (europa.eu)). In a copyright suit, it would be highly relevant to know how much certain copyrighted material was considered by the AI, whether it was copied directly or indirectly, and the weight given to certain source material over other source material.

Another issue that could arise is fair use. There are significant and complex questions in both the copyright area generally and in the AI space specifically as to when a newly created work is considered to be a "transformative" fair use and allowable versus when a newly created work is considered to be a "derivative work" and copyright infringement. The test for fair use is a multifactor test, but there is substantial case law interpreting the factors and discussing their weight. See 4 Nimmer on Copyright Section 13.05 (2022)("The statute gives no guidance as to the relative weight to be ascribed to each of the listed factors."). The first factor, "the purpose and character of the use," would explore whether the AI is being used as reference material or whether there is a for profit motive behind the use. At this point, many of the AI tools out there are available for free, but the motive may be to capture market share and eventual profit. The second factor, "the nature of the work," will certainly be important because training material can range the gamut from highly creative artistic works to what might be considered routine software code for common functions that does not have much room for creativity (while it should be noted some software code can also be highly creative and customized). The third factor, "the amount and substantially of the portion used," likely will be very important in considering whether the use by the AI was fair, but this might be difficult to show if the AI is not required to be transparent about its process in a traceable and explainable matter. Finally, the effect that the AI generated works "has on the market for the original works" will have to be considered. The availability of an AI generated work inspired by the

work of Andy Warhol would seem to diminish the value of the original as would AI generated software code that solves the same problem as the code that was used to train the AI. (Cf. https://www.reuters.com/world/us/us-supreme-court-hears-arguments-andy-warhol-copyright-dispute-2022-10-12/)(discussing whether an Andy Warhol piece that was based on a photograph of Prince was sufficiently transformative to benefit from a fair use defense). We suspect that AI generated works might be subject to higher scrutiny because there is a bias that humans can be independently creative while a programmed machine is just doing rote steps that lack the spark of independence.

In cases involving human creators, a judge and jury can hear from the creator about what they did, what inspired them and the process they used to make an allegedly infringing work. In a case where AI is using training material to create new works, the ultimate creator, the AI, cannot testify. Even if it is fully transparent what the AI did to create the work, the factual and legal questions surrounding copying, independent creation and fair use will not be simple to sort out.

Joseph Casino, *a partner with Wiggin & Dana, is the leader of the firm's intellectual property practice group, he advises clients on patent office disputes, complex licensing, patent monetization, patent portfolio development, and worldwide intellectual property strategic matters.*

Michael J. Kasdan, *a partner with the firm, focuses on all areas of intellectual property law, providing his clients with full-service IP expertise that ranges from patent, trademark, copyright and trade secret litigation to IP-related transactions.*