Q&A: the crossroads where generative AI meets copyright law

OCTOBER 9, 2023

Westlaw Today asked Robins Kaplan LLP attorney Bryan J. Mechell to provide some guidance to understanding the many copyright controversies that have accompanied the introduction of generative artificial intelligence systems.

Westlaw Today: Generative AI has captivated the world in the past year. In the U.S., its introduction has been plagued with legal disputes and speculation. This presents challenges for companies protecting their generative AI innovations as well as for users understanding rights and risks associated with generative AI tools. With regard to AI and intellectual property, there have been two primary copyright questions that everyone is asking. The first is: Can something created with AI be protected by copyright law?

Bryan Mechell: In short, yes, content created using a generative AI tool can likely be protected by copyright law — but the scope of how much human input is necessary to qualify the user of an AI system as an "author" of the generated work is still an open question subject to substantial ongoing legal and regulatory discussion.

While generative AI models that learn from datasets as large as the internet can be exceptionally powerful, those datasets are heavily interspersed with copyrighted and other protected material.

The U.S. District Court for the District of Columbia recently considered aspects of this question in *Thaler v. Perlmutter*, No. 22-cv-1564, 2023 WL 5333236 (D.D.C. Aug. 18, 2023). In that case, the court affirmed the U.S. Copyright Office's denial of an application for an AI-generated image that was generated autonomously by an AI system called the "Creativity Machine."

Noting that "human authorship is an essential part of a valid copyright claim," the court highlighted Section 102(a) of the Copyright Act, 17 U.S.C.A. § 102(a), which provides copyright protection to "original works of authorship fixed in any tangible medium of expression, now known or later developed, from which they can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device."

Section 101 of the act, 17 U.S.C.A. § 101, further provides that the work of authorship must be fixed "by or under the authority of the author." This "authorship" requirement, the court noted, was "presumptively human" and centered on "acts of human creativity." The court noted that "copyright has never stretched so far, however, as to protect works generated by new forms of technology operating absent any guiding human hand."

Notably, however, the *Thaler* decision left the critical and more fact-specific question unanswered of how much human input would have been needed to qualify the work for protection. While courts have long recognized that technological tools can be used by authors as part of the creative process, generative AI highlights important questions about how a technological tool can be used by a human author and the extent of human decision-making required. With the right amount of human input and creativity, it stands to reason that works containing outputs from advanced technological tools may qualify for copyright protection. Courts and the U.S. Copyright Office are likely to provide useful guidance as they explore the contours of this issue in the coming year.

WT: The second question concerns whether generative AI companies such as OpenAI are violating copyright law, as some class actions have been filed recently over infringement and related issues. Who is waging these suits and what are these plaintiffs claiming?

BM: Proposed class-action lawsuits filed this year against GitHub, Stability AI, OpenAI and Meta — including recent actions filed by George R.R. Martin, John Grisham, Pulitzer Prize winner Michael Chabon, comedian and author Sarah Silverman, and various other authors against OpenAI and Meta — raise important questions about liability for unauthorized use of copyrighted materials to train generative AI models, as well as questions about ownership of generative AI outputs.

These actions include allegations that generative AI companies trained their generative AI tools on protected materials without proper attribution or compensation. For example, the class-action complaint filed against GitHub, Microsoft, OpenAI and related corporate groups in November 2022 alleges that the defendants trained Codex and Copilot (coder-assisting generative AI programs) on public code that was protected by open-source licenses, but

THOMSON REUTERS®

the AI does not provide attribution of authorship or copyright when outputting that code. These are alleged Digital Millennium Copyright Act, or DMCA, violations.

The various class-action litigations filed against OpenAI and Meta allege that the generative AI tool uses copyrighted works in its vast training datasets that are built by scraping the internet for text data — which necessarily leads the tool to capture, download and copy copyrighted written works, plays and articles. The complaints also assert that the outputs of the generative AI model — i.e., the text-generated responses to a user input query — constitute copyright infringement.

For intellectual property owners protecting their generative AI innovations, as well as end users who have obtained licenses to use generative AI tools, these lawsuits underscore the importance of closely monitoring the composition of generative AI training datasets, scope and content of outputs, and license terms regulating the use of these rapidly evolving technologies.

WT: What are the defendants claiming gives them the right to use copyrighted content to train their systems?

BM: OpenAI has moved to dismiss the bulk of the claims in the class action filed by Sarah Silverman and others — the "heart" of which it argues are copyright claims — on the basis that they "misconceive the scope of copyright, failing to take into account the limitations and exceptions (including fair use) that properly leave room for innovations like the large language models now at the forefront of artificial intelligence." *Silverman v. OpenAI Inc.*, No. 23-cv-3417, *motion to dismiss filed* (N.D. Cal. Aug. 28, 2023).

The Copyright Act grants a limited monopoly in service of a broader goal to — as the U.S. Constitution states — "promote the Progress of Science and useful Arts." But this protection has limits, including the "fair use" doctrine, which OpenAI argues should be adapted to account for "rapid technological change" and, in short, to protect the use of large sets of training data for generative AI models. OpenAI argues in its motion to dismiss that current judicial precedent supports the conclusion that it is not an infringement to create "wholesale cop[ies] of [a work] as a preliminary step" to develop a new, non-infringing product, even if the new product competes with the original. *Google LLC v. Oracle America Inc.*, 141 S. Ct. 1183 (2021) (summarizing *Sega Enters. Ltd. v. Accolade Inc.*, 977 F.2d 1510 (9th Cir. 1993)); *see also Sony Comput. Ent. v. Connectix Corp.*, 203 F.3d 596 (9th Cir. 2000).

WT: You mention the potential liability of those training their AI models, but how are technology companies addressing the risk of developing and using generative AI models?

BM: Technology and software license disputes involving intellectual property and contract rights carry significant risk in terms of potential business disruption and damages. While generative AI models that learn from datasets as large as the internet can be exceptionally powerful, those datasets are heavily interspersed with copyrighted and other protected material. The increasing implementation and use of generative AI at software and technology companies could, therefore, lead to increased disputes over the use of copyrighted data to train generative AI models as well as ownership of outputs.

It is likely going to be some time before we get solid guidance from courts, regulations and potentially Congress on the scope of various IP rights in generative AI tools. In the meantime, it is critical that technology companies developing and licensing generative AI innovations closely monitor, catalog and assess training data used by generative AI tools. This includes maintaining a detailed record of the sources, libraries and metadata, and the compositions of each — which provides the basic materials needed to assess risks associated with an AI system trained on protected materials.

All aspects of the generative AI ecosystem are important to consider from a risk management perspective, including the training set, the AI algorithm or model itself, the input query, and the output result. One strategy is to develop a cross-functional team tasked with monitoring use and compliance. As part of this assessment, companies should pay close attention to license terms that outline authorized uses and protect IP rights, assess how generative AI outputs are being used (and modify licenses accordingly), and develop a robust review process for monitoring compliance with developing laws and regulations.

WT: What are some strategies for crafting effective license terms in software license agreements to maximize benefits of IP protection for generative AI innovations?

BM: One important takeaway for technology companies leveraging generative AI innovations is to take a holistic approach to licensing that acknowledges how any generative AI tools interact with licensed software. This includes drafting terms that clearly articulate what rights are licensed, authorized uses, restrictions and warranties — all of which can vary based on the specific piece of the generative AI ecosystem under consideration. For example, license agreements should identify the scope and content of training data used by any generative AI tools, and how (or if) user data is used to train the model. Similarly, license agreements should define ownership and authorized uses of the generative AI outputs and articulate restrictions on how the overall tool can be used. It is important to remain mindful of internal goals for IP protection in generative AI and implement intentional processes for refining licensing practices as the laws applicable to generative AI evolve.

About the author

Bryan J. Mechell is a trial attorney in the intellectual property and technology group at **Robins Kaplan LLP** in Minneapolis, where he focuses his practice on litigating complex technology and software license disputes. Drawing on his background in computer science and physics, Mechell counsels emerging and established businesses that are developing and leveraging cutting-edge technologies, and he helps them develop methods for protecting innovation, mitigating risk and enforcing license compliance. He can be reached at bmechell@robinskaplan.com.

This article was first published on Westlaw Today on October 9, 2023.