SPRING 2023 E-NEWSLETTER

At Digital Mountain, we assist our clients with their electronic discovery, digital forensics, cybersecurity, and data analytics needs. For this E-Newsletter, we discuss ChatGPT and AI innovations, as well as legal and data security implications.

Do ChatGPT and Similar Technologies Violate IP Law?

ChatGPT, Bard, and others, are churning out material that's showing up on an increasingly pervasive scale wherever we find written content. While generating wholly AI-created content sounds like a monumental leap forward, we may need to bookmark this moment and ask some important questions, especially when it comes to confidential or protected data. Take, for example, the Samsung employees who used ChatGPT to help with their work, only to discover that they were releasing confidential company information (https://www.techradar.com/news/samsung-workers-leaked-company-secrets-by-using-chatgpt). For those interested in Intellectual Property matters, the rise of LLM content creation raises an essential question: Do these tools violate IP law?

The short answer is yes, probably easily. In fact, we'd say the hypothetical likelihood is very high that responses generated by ChatGPT, Bard, and others borrow substantially from copyrighted material. The training datasets for LLMs are built from massive document sets with almost infinite sources. When we asked Chat-GPT about copyrighted material in its training dataset, it confirmed that, "…I was trained on large datasets of text-based information, including a wide range of materials such as books, articles, and other written content that may be subject to copyright protection," (https://chat.openai.com/chat accessed April 7, 2023 by author).

The concern that LLMs might infringe upon IP protections also applies to software code. In November 2022, the same month ChatGPT was released to the public, a class action suit was filed against online open-source code platform GitHub, OpenAI (the company that released ChatGPT), and their parent company, Microsoft, claiming that "GitHub and OpenAI launched Copilot, an AI-based product that promises to assist software coders by providing or filling in blocks of code using AI." According to the complaint, integral to Copilot is Open AI's Codex. Codex is a language model which helps programmers and coders with coding tasks. Codex was trained on a large dataset of software code and coding language. The lawsuit alleges that Copilot removes the licensing information while making use of the data, hence, violating intellectual property

protections. The case, *J. Doe 1, et al. v. GitHub, Inc., et al.*, US Dist Ct., NDCA, Case 3:22-cv-06823, (2022), is certainly going to be one to watch as the defendants, GitHub, and its parent company Microsoft, as well as sibling OpenAI, have billions of dollars riding on where intellectual property protections lie.

Another lawsuit filed by the same attorneys who filed *Doe v GitHub*, alleges that AI image generator, Stable Diffusion, and its parent company, Stability AI, violated the intellectual property protections of visual/graphic artists by including unlicensed, copyrighted works of art in its training dataset. In *Sarah Anderson, et al v. Stability AI, Ltd, et al,* US Dist. Ct., NDCA Case 3:23-cv-00201, (2023), plaintiffs claim that Stable Diffusion is "a complex collage tool" pulling together bits and pieces of images from its training dataset, which includes images for which a licensing fee should have been paid to the original artists. Notably, in October 2022, Open AI and Shutterstock announced that to train DALL-E, another AI image generator, Open AI paid to license Shutterstock's catalog, and in return, Shutterstock would be offering AI-created images on its AI platform. In addition, Shutterstock announced the creation of a fund and framework to ensure artists were compensated for their images used in AI image creation (https://www.prnewswire.com/news-releases/shutterstock-partners-with-openai-and-leads-the-way-to-bring-ai-generated-content-to-all-301658310.html). Does that sound like an admission that they knew they should have been paying license fees?

Whether the old adage, "Everything is derivative," is true or false, we know for sure that there's an abundance of creative work in the world that can be used for training. While the goal of the developers may be to create a better tool, lawyers will be trying to protect the rights of artists, authors, coders, and creators. Other lawyers might try to argue some angle of the educational use exemption to copyright protection. That would be a creative piece of work on its own. Either way, we imagine we will be doing some fascinating eDiscovery in the process of sorting all this out.

Please direct questions and inquiries about electronic discovery, digital forensics, cybersecurity, and data analytics to info@digitalmountain.com.

UPCOMING INDUSTRY EVENTS

RSA CONFERENCE 2023
San Francisco, CA: April 24-27, 2023

MASTER'S CONFERENCE APRIL 2023
Chicago, IL: April 25, 2023

SEDONA CONFERENCE WORKING GROUP 1 MIDYEAR MEETING 2023
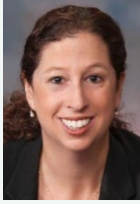Portland, OR: April 27-28, 2023

IP COUNSEL CAFE ANNUAL MEETING
Palo Alto, CA: May 3-4, 2023

SEDONA CONFERENCE WORKING GROUP 11 ANNUAL MEETING 2023
Denver, CO: May 4-5, 2023

Click here to see more upcoming events and links.

Digital Mountain, Inc. Founder and CEO, Julie Lewis, will be presenting at various upcoming industry events. Please send requests for speaker or panel participation for her to marketing@digitalmountain.com.

DIGITAL MOUNTAIN, INC.
4633 Old Ironsides Drive, Suite 401
Santa Clara, CA 95054
866.DIG.DOCS

www.digitalmountain.com

Contact us today!

FOLLOW US AT: