Copyright Infringement Lawsuit Alleges OpenAI's Unauthorized Use of US Authors' Content to Train AI Chatbot

Two U.S. authors have filed a proposed class action lawsuit against OpenAI, a leading artificial intelligence (AI) company, alleging copyright infringement. The San Francisco federal court lawsuit accuses OpenAI of misusing the authors' works to train its popular generative AI system, ChatGPT.

The plaintiffs in this case are Massachusetts-based writers Paul Tremblay and Mona Awad. They claim that ChatGPT utilized data copied from thousands of books without obtaining proper permissions, thereby infringing upon the copyrights of the authors. Tremblay and Awad argue that OpenAI's actions amount to unauthorized use of their creative content.

Representatives for the authors declined to provide further comments, while OpenAI, a private company with Microsoft Corp as its backer, has not yet responded to the allegations made in the lawsuit.

This lawsuit is just one of several legal challenges that have emerged concerning the use of copyrighted material in training advanced AI systems. Other cases involve disputes between source-code owners and OpenAI as well as Microsoft's GitHub, along with visual artists suing companies like Stability AI, Midjourney, and DeviantArt.

Ready to make a change in your legal career? BCG Attorney Search can help you find the perfect job.

The defendants in these cases have consistently argued that their AI systems employ fair use principles when utilizing copyrighted works. ChatGPT, specifically, is an AI system designed to respond to text prompts in a conversational manner. Since its launch, it has gained immense popularity and quickly became the fastest-growing consumer application in history, reaching an impressive 100 million active users within just two months.

Generative AI systems like ChatGPT rely on extensive amounts of data scraped from the internet to create content. According to the lawsuit filed by Tremblay and Awad, books play a vital role in this process as they offer examples of high-quality long-form writing, which is essential for training the AI system effectively.

The plaintiffs estimate that OpenAI's training data comprises over 300,000 books, including content sourced from illicit "shadow libraries" that distribute copyrighted books without permission.

Mona Awad is widely recognized for her novels, including "13 Ways of Looking at a Fat Girl" and "Bunny," while Paul Tremblay's notable works include "The Cabin at the End of the World," which was adapted into the M. Night Shyamalan film titled "Knock at the Cabin," released in February.

To support their claim, Tremblay and Awad contend that ChatGPT is capable of generating highly accurate summaries of their books, indicating that their works were indeed present in the AI system's database.

The lawsuit seeks unspecified monetary damages on behalf of a nationwide class of copyright owners whose works were allegedly misused by OpenAI. The case highlights the significance of protecting intellectual property rights in the rapidly evolving field of AI, where the boundaries of fair use and unauthorized utilization of copyrighted material are being tested.

As this legal battle unfolds, it will undoubtedly contribute to shaping the future landscape of AI development and its relationship with intellectual property. The outcome of this lawsuit may set important precedents for the responsible and ethical use of copyrighted works in training AI models and could impact the practices of other AI companies in the industry.

Don't be a silent ninja! Let us know your thoughts in the comment section below.