

Microsoft asks to dismiss New York Times's 'doomsday' copyright lawsuit

Publication Date: 2024-03-07

Author: Edward Helmore

Section: Technology

Tags: Technology, New York Times, Microsoft, Artificial intelligence (AI), OpenAI, Computing, Newspapers, news

Article URL: <https://www.theguardian.com/technology/2024/mar/06/microsoft-dismiss-ai-new-york-times-lawsuit>



Microsoft has responded to a copyright infringement lawsuit brought by the New York Times over alleged use of content to train generative artificial intelligence, calling the claim a false narrative of “doomsday futurology”. The tech giant said the lawsuit was near-sighted and akin to Hollywood’s losing backlash against the VCR. In a motion to dismiss part of the lawsuit filed on Monday, Microsoft, which was sued in December alongside ChatGPT-maker OpenAI, scoffed at the newspaper’s claim that Times content receives “particular emphasis” and that tech companies “seek to free-ride on the Times’s massive investment in its journalism”. In the lawsuit – which could have major implications for the future of generative artificial intelligence and for news-content production – the Times alleged that Microsoft, which is OpenAI’s biggest investor, had unlawfully used the paper’s “copyrighted news articles, in-depth investigations, opinion pieces, reviews, how-to guides, and more” to create artificial intelligence products that “threatens The Times’s ability to provide that service”. But in its response, Microsoft said the lawsuit was akin to Hollywood’s resistance to the VCR that consumers used to record TV shows and which the entertainment business in the late 1970s feared would destroy its economic model. “The VCR is to the American film producer and the American public as the Boston strangler is to the woman home alone,” Microsoft said in its response, quoting from congressional testimony delivered by Jack Valenti, then head of the motion picture association of America, in 1982. In this case, Microsoft said, the Times was attempting to use “its might and its megaphone to challenge the latest profound technological advance: the Large Language Model.” Microsoft’s lawyers also argued that “content used to train LLMs does not supplant the market for the works, it teaches the models language”. OpenAI has already asked a judge to dismiss parts of the lawsuit against it, alleging that the publisher “paid someone to hack OpenAI’s products” to create examples of copyright infringement using its ChatGPT. “ChatGPT is not in any way a substitute for a subscription to The New York Times,” attorneys for OpenAI wrote. “In the real world, people do not use ChatGPT or any other OpenAI product for that purpose. Nor could they. In the ordinary course, one cannot use ChatGPT to serve up Times articles at will.” But the Times hit back after Microsoft filed its legal response, again taking issue with the analogy of 1980’s home-taping technology. “Microsoft doesn’t dispute that it worked with OpenAI to copy millions of The Times’s works without its permission to build its tools,” said Ian Crosby, lead counsel for the New York Times, in an emailed response, adding that Microsoft “oddly compares LLMs to the VCR even though VCR makers never argued that it was necessary to engage in massive copyright infringement to build their products”. The statement continued: “Despite Microsoft’s attempts to frame its relationship with OpenAI as a mere ‘collaboration,’ in reality, as The Times’s complaint states, the two companies are intertwined when it comes to building their generative AI tools.” But the battle comes against a series of lawsuits brought by authors and artists over various

aspects of copyright, including ownership of creative work created using the technology, and complaints that AI technology can create some wildly misleading information, what the industry cutely calls “hallucinations”. Last month, Google was forced to apologize when its Gemini chatbot was used to create images of Black soldiers in second world war-style German military uniforms and Vikings in traditional Native American dress. Google temporarily suspended the technology’s ability to make images of people and vowed to fix what it described as “inaccuracies in some historical” depictions. The twin concerns – that AI technology can violate copyrighted material and create information or images that are wildly improbable – comes as OpenAI recently acknowledged that it was “impossible” to train AI models without copyrighted works “because copyright today covers virtually every sort of human expression”. OpenAI has refused to disclose the contents of its training databases, including for its newest tool, a video generator called Sora. In a letter to the UK’s House of Lords, the company also said that “limiting” the training data to content in the public domain “would not provide AI systems that meet the needs of today’s citizens”. The OpenAI CEO, Sam Altman, said in January that he was “surprised” by the Times’s lawsuit because the system did not need the Times data to train itself. “I think this is something that people don’t understand. Any one particular training source, it doesn’t move the needle for us that much.” Altman claimed the Times’s articles represented a minute part of the corpus of text used to create ChatGPT.