

NYT vs OpenAI, Microsoft: The debate on AI, copyright

SOUMYARENDRA BARIK
NEW DELHI, DECEMBER 28

THE NEW York Times (NYT) has sued OpenAI – the firm behind generative artificial intelligence (AI) platform ChatGPT – and Microsoft for “unlawful” use of the paper’s copyrighted content. The lawsuit could frame the legal contours around intellectual property (IP) rights in the age of generative AI platforms, which scrape the Internet to build and train their large language models.

NYT is the first major news publisher to sue the companies over their generative AI offerings. “Defendants seek to free-ride on The Times’s massive investment in its journalism,” the complaint says, accusing OpenAI and Microsoft of using its content “without payment to create products that substitute for The Times and steal audiences away from it”.

The issue is symbolic of the larger debate on how generative AI platforms could affect people from the creative industry given such systems are built on the back of work done by creators of original content, which is then synthesised through an algorithm and presented as fresh information by the AI systems. Earlier this year, two US authors

had also sued OpenAI, claiming in a proposed class action that the company misused their works to “train” ChatGPT.

NYT’s contention

The lawsuit contends that millions of articles published by publication were used to train automated chatbots, which now compete with the news outlet as a source of reliable information. Microsoft has a sizable investment in OpenAI.

NYT reported that it approached Microsoft and OpenAI in April to raise concerns about the use of its intellectual property. But the talks had not produced a resolution, the publication said.

The publication has also alleged OpenAI and Microsoft’s large language models (LLMs), which power ChatGPT and Copilot, “can generate output that recites Times content verbatim, closely summarises it, and mimics its expressive style.” This “undermine[s] and damage[s]” the Times’ relationship with readers, while also depriving it of “subscription, licensing, advertising, and affiliate revenue”.

In August, NYT had blocked OpenAI’s web crawler.

AI and IP rights

Ever since generative AI platforms like ChatGPT and Google’s

Bard have burst onto the scene, it has ignited a debate over intellectual property rights over original content on the Internet.

Such AI platforms typically scrape through the Internet to gather data to feed into their large language models after which they tune it to produce results. The music business, too, is pushing back on the use of AI in the industry. The Universal Music Group, for instance, has been asking music streaming services like Spotify to stop developers from scraping its material to train AI bots to make new songs.

The debate is gaining traction at a time when countries around the world, including India, have archaic copyright laws that need reimagining keeping the AI wave in mind. For instance, In India, creative works are regulated under the Copyright Act of 1957.

The definition of an “author” under the Act includes any literary, dramatic, musical or artistic work which is computer generated, the person who causes the work to be created. But, that definition does not take into account that AI systems do not generate information on their own, they are simply as good as the base dataset on which they are trained. And the base dataset is made by copyrighted work produced by other authors.

FULL REPORT ON
www.indianexpress.com

EXPLAINED

E.