

IP Laws - ChatGPT vs. NY Times

Arnaud Capitan

Paul Archer

Julien Waxweiler

Axel Guimbault

Louis Giordmaina

Kenzo Cao Van

Clémence Audibert

Noé Mosca

Pierre-Antoine Palmero

Damien Pelissier

Annabelle Montaudon

Eric Romagnan

• In what ways does GPT-4 use New York Times articles?

The training of GPT-4 and LLM in general is based on the training of these models with billions of data scraped from the Internet. In order to assure the best performance, OpenAI recognized that it needed “good-quality” data such as the ones from the Times since it is well-spoken English. The Times is one of the main sources of the huge amount of training data, arriving in third place just behind Wikipedia. This causes an issue since the goal of the LLMs is to statistically answer to a prompt by producing the most probable words in a row, and therefore mimicking the Times style and content. It threatens the work of the Times since the information is not reliable anymore.

GPT-4 uses New York Times articles : points 1 to 3 of the complaint. GPT is trying to free ride on the Time’s investment on journalism. Microsoft Bing is also one of the target of the complaint.

- Microsoft and OpenAI are going to copy further the NYTimes’ articles to create GPT5 (art. 96)

In addition, OpenAI and Microsoft cooperated in order to recreate synthetic data to provide for their LLMs and ensure continuing training. Times regrets that this pure mimicking by reproducing content and style impacts the work of Times, since it is not recognized by the defendants and is not known by the public.

“They gave Time’s particular emphasis when building their LLMs, revealing a preference that recognize the value of those works” art.3 Not only did OpenAI use Time’s data, it is one of the most used source of data to mine for them, because of Time’s originality, quality and precision in their data. But as we saw, originality is the prime aspect of Intellectual Property (IP). OpenAI use of Time’s data **cannot** be used in the case of Fair use because Time’s will lose exclusivity on their data and way of doing articles. They directly enter in competition with OpenAI regarding news and information worldwide. Especially when Time’s style is imitated, the effect on the potential market of Times is huge.

• Why does New York Times assert that there is copyright infringement?

- In support of the New York Times, Open AI’s practice is in fact an infringement on the newspapers’ copyrights

They copy the whole content of articles without any transformation, which falls under the description of copyright infringement (see the image below).

In the context of Open AI using information from New York Times, Fair use is described as follows :

- + Nature of the copyrighted work here falls under the Fair Use exception, as the dissemination of facts or information benefits the public.
 - You can use parts of a work, but by using everything you depreciate the original work.
 - The purpose and character of your use : you need to transform the data you're using, which here is clearly not the case, as GTP4 uses the article from The New York Times as its own.
 - The effect of the use upon the potential market, which will be described later on
-
- They have no license (article 97)
 - In this case, the vast quantity of articles utilized suggests that the usage was not measured or limited, potentially exceeding what would be considered fair use.

Exhibit J at 5.

100. Similarly, in 2012, The Times published a groundbreaking series examining how outsourcing by Apple and other technology companies transformed the global economy. The series was the product of an enormous effort across three continents. Reporting this story was especially challenging because The Times was repeatedly denied both interviews and access. The Times contacted hundreds of current and former Apple executives, and ultimately secured information from more than six dozen Apple insiders. Again, GPT-4 copied this content and can recite large portions of it verbatim:²⁷

Output from GPT-4:

many of America's other global companies — aren't nearly as avid in creating American jobs as other famous companies were in their heydays. Apple employs 43,000 people in the United States and 20,000 overseas, a small fraction of the over 400,000 American workers at General Motors in the 1950s, or the hundreds of thousands at General Electric in the 1980s. Many more people work for Apple's contractors: an additional 700,000 people engineer, build and assemble iPads, iPhones and Apple's other products. But almost none of them work in the United States. Instead, they work for foreign companies in Asia, Europe and elsewhere, at factories that almost all electronics designers rely upon to build their wares. "Apple's an example of why it's so hard to create middle-class jobs in the U.S. now," said Jared Bernstein, who until last year was an economic adviser to the White House. "If it's the pinnacle of capitalism, we should be worried." Apple executives say that going overseas, at this point, is their only option. One former executive described how the company relied upon a Chinese factory to revamp iPhone manufacturing just weeks before the device was due on shelves. Apple had redesigned the iPhone's screen at the last minute, forcing an assembly line overhaul. New screens began arriving at the plant near midnight. A foreman immediately roused 8,000 workers inside

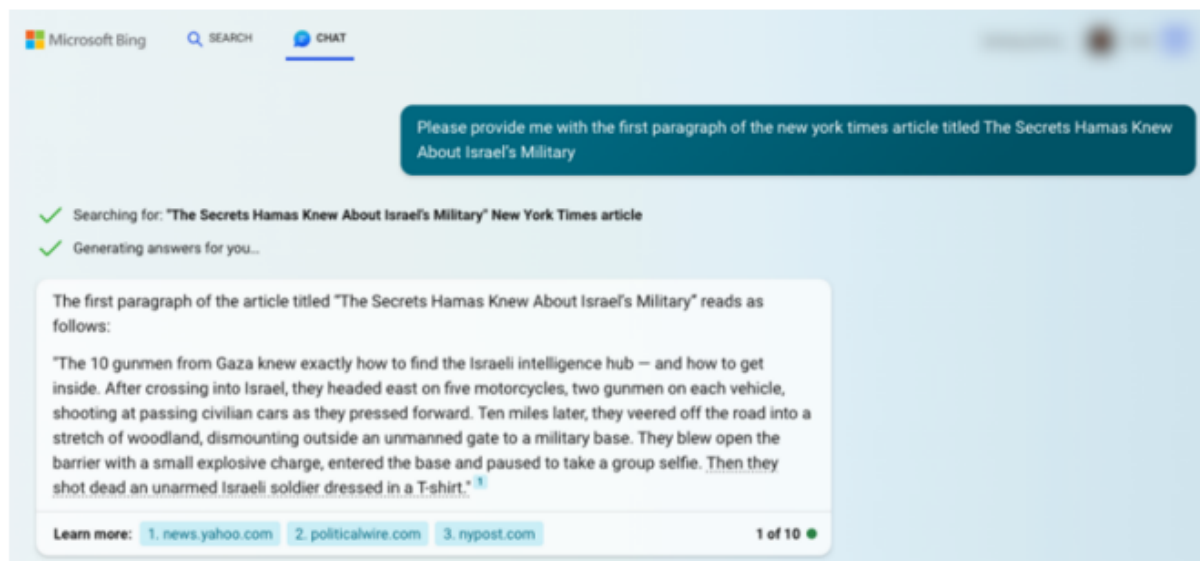
Actual text from NYTimes:

many of its high-technology peers — are not nearly as avid in creating American jobs as other famous companies were in their heydays. Apple employs 43,000 people in the United States and 20,000 overseas, a small fraction of the over 400,000 American workers at General Motors in the 1950s, or the hundreds of thousands at General Electric in the 1980s. Many more people work for Apple's contractors: an additional 700,000 people engineer, build and assemble iPads, iPhones and Apple's other products. But almost none of them work in the United States. Instead, they work for foreign companies in Asia, Europe and elsewhere, at factories that almost all electronics designers rely upon to build their wares. "Apple's an example of why it's so hard to create middle-class jobs in the U.S. now," said Jared Bernstein, who until last year was an economic adviser to the White House. "If it's the pinnacle of capitalism, we should be worried." Apple executives say that going overseas, at this point, is their only option. One former executive described how the company relied upon a Chinese factory to revamp iPhone manufacturing just weeks before the device was due on shelves. Apple had redesigned the iPhone's screen at the last minute, forcing an assembly line overhaul. New screens began arriving at the plant near midnight.

110: If the data has already been read on ChatGPT based on the articles from the New York Times, there is no need to check the original source.

111: Explain the examples that clearly indicate that ChatGPT is not using the New York Times articles under the Fair Use exception. (112-)

112: Pretty self-explanatory, it just gives the article bypassing the payroll.



Build a legal argument In support of the New York Times (argue that Open AI's practices infringe on NYT copyright)

Legal Argument in Favor of The New York Times Against OpenAI

1. Unauthorized use of New York Times' works

The Copyright Act grants authors from The New York Times exclusive rights to their works. Any reproduction without authorization constitutes infringement and does not fall under the fair use exception.

OpenAI used millions of New York Times articles to train its AI models without obtaining a license or permission, even when the articles were protected by paywall.

By reproducing them without any change, ChatGPT can generate substantial portions of NYT articles, sometimes word-for-word, proving unauthorized use, again not falling under the fair use exception, as it would require transformation of data.

OpenAI profits from AI models trained on NYT content without compensating the publisher, thus requiring a fee, or at least a percentage of the revenue from the use of NYT data..

2. Harm to NYT's Market

ChatGPT use of NYT articles led to revenue loss for the NYT, as the newspaper relies on subscriptions and digital advertising, both of which are impacted when AI-generated content substitutes for its original articles. Users of ChatGPT can get freely the articles protected by paywall from the New York Times, which is clearly an issue as it removes a revenue stream.

It also led to reduced website traffic. AI tools provide detailed summaries and excerpts, discouraging users from visiting NYT's website as they can get all the contents already from ChatGPT.

It constitutes unfair competition, as OpenAI benefits financially from NYT's journalism without contributing to its costs, creating a free-riding business model that undermines investigative journalism.

3. Fair Use Defense is Inapplicable

OpenAI claims its use is "transformative," but ChatGPT reproduces entire excerpts rather than creating new, original works.

Summarizing or paraphrasing does not constitute a sufficient transformation when the purpose remains the same—delivering news content.

Maybe more importantly, fair use does not apply if the use harms the market of the original work. NYT experiences direct financial losses from AI-generated content replacing visits to its site.

And to characterize fair use, one must take into account the nature of the work. NYT's journalism is highly original and creative, making it more strongly protected under copyright law. Therefore fair use should not and could not be invoked in this case by the defendant company, OpenAI.

4. Misinformation and Reputational Harm

One of the most worrisome elements for the New York Times is the reputational harm ChatGPT can cause and may have already caused for the news organisation. Indeed ChatGPT has generated fake or inaccurate quotes and articles, falsely attributing them to NYT, and on the contrary claiming New York Times articles' contents as its own data.

And this may have very long-term impacts. If misinformation spreads under NYT's name, the journal's reputation for accuracy and trustworthiness is compromised.

Conclusion

OpenAI's mass copyright infringement and commercial exploitation of NYT's content violate U.S. copyright law.

The fair use defense is invalid due to lack of transformation and significant market harm.

AI-generated misinformation attributed to NYT poses additional risks to journalistic integrity.

Legal action is necessary to protect NYT's rights and ensure fair compensation for its work.