# News publishers push back as AI firms face scrutiny over content use



The growing conflict between AI developers and online news publishers over the use of copyrighted content is reaching a critical point, with far-reaching implications for journalism, the internet and democratic society. At issue is how generative AI systems rely on journalistic content to function, even as they increasingly divert traffic and revenue from original news sources.

Large language models depend heavily on human-generated journalism to deliver accurate and context-rich outputs. News content not only grounds these systems in reality but also helps mitigate bias and misinformation. Yet publishers argue that AI features such as zero-click search responses and retrieval augmented generation are eroding web traffic and advertising income by serving users answers without directing them to the original source.

This concern is exemplified by Google’s AI Overviews, which summarise information directly in search results. Publishers report a decline in referral traffic since its introduction, with news brands particularly affected. As AI tools adopt advertising-driven business models, publishers fear further revenue loss without fair attribution or compensation.

Meanwhile, AI-driven bots and crawlers are adding operational burdens by harvesting content without providing equivalent returns. These bots often ignore protocols like robots.txt, and attempts to block them risk harming search visibility altogether.

Legal responses are starting to emerge. In the US, a judge recently ruled that Anthropic’s AI training on copyrighted books may constitute fair use, though the company still faces trial over alleged use of pirated material. In Europe, the French competition authority fined Google €250 million for training its chatbot on publisher content without proper consent.

Some AI firms are pursuing licensing deals. Meta is in talks with publishers including News Corp and Axel Springer, while OpenAI and Amazon have also secured agreements. In Japan, many news organisations have banned AI companies from using their content, but unauthorised scraping continues, raising legal and ethical concerns.

The issue is complicated by journalism’s own reliance on fair use for commentary, criticism and historical analysis. As regulatory bodies such as the US Copyright Office revisit their guidance, publishers and policymakers are seeking a middle ground—protecting content creators while allowing for responsible AI innovation.

Proposals include transparent licensing models, improved bot management and safeguards against monopolistic practices. Both sides recognise the need for a functioning digital economy that supports high-quality journalism and open access to information.

The relationship between AI firms and publishers remains tense yet interdependent. Without collaborative solutions, the economic foundation of professional journalism could weaken. But with thoughtful policy and mutual engagement, there remains a path to ensure that AI advances in tandem with a free and sustainable press.

Created by [Amplify](https://www.hbmadvisory.com/amplify): AI-augmented, human-curated content.

## Bibliography

1. <https://publicknowledge.org/is-there-a-middle-ground-in-the-tug-of-war-between-news-publishers-and-ai-firms-part-1-framing-the-problem/> - Please view link - unable to able to access data
2. <https://www.reuters.com/business/meta-talks-with-axel-springer-fox-others-ai-news-licensing-wsj-reports-2025-09-18/> - Meta Platforms has engaged in discussions with media companies such as Axel Springer, Fox Corp, and News Corp to license their news content for integration into Meta's AI products. These negotiations aim to secure rights to use news and other media in AI applications, aligning with Meta's broader effort to strengthen its AI division. This move mirrors similar partnerships formed by other AI companies, such as OpenAI and Perplexity, which have also sought deals with news outlets. Notably, Reuters signed a licensing deal with Meta in 2024, and Amazon partnered with The New York Times for AI content integration in May. Axel Springer declined to comment on the talks, and Meta, Fox, and News Corp have not responded publicly.
3. <https://www.reuters.com/technology/french-competition-watchdog-hits-google-with-250-mln-euro-fine-2024-03-20/> - The French competition watchdog fined Google 250 million euros for breaching EU intellectual property regulations regarding its AI service, specifically the AI-powered chatbot Bard, now rebranded as Gemini. Bard was trained using content from publishers and news agencies without notifying them. Google, which will not contest the facts, agreed to this as part of settlement proceedings and proposed a series of remedies. They considered the fine disproportionate, claiming their efforts were not sufficiently acknowledged and accepted the settlement to focus on sustainable practices and better relations with French publishers. The issue originated from complaints by major French news organizations, resolved in 2022 with Google agreeing to certain commitments. The watchdog cited that Google violated several settlement terms and hindered fair negotiation practices by linking AI content use to protected content display. This fine underscores a broader concern among publishers and writers about AI services extracting their content without consent or proper compensation.
4. <https://apnews.com/article/1e5cece51c2e4bd0bb21d94de2abb035> - A federal judge has ruled that AI company Anthropic did not violate U.S. copyright law by training its Claude chatbot on millions of copyrighted books, deeming the use 'quintessentially transformative' and therefore fair use. However, Anthropic must face trial in December over allegations it illegally downloaded those books from pirated online libraries. The lawsuit, brought by authors Andrea Bartz, Charles Graeber, and Kirk Wallace Johnson, accuses Anthropic of large-scale theft to profit from the intellectual labor in those books. Internal documents showed employee concerns over the legality of using pirated sources. Although Anthropic later began purchasing and digitizing books more lawfully, the judge stated that prior piracy could still incur liability. The outcome of this case may influence similar lawsuits against other AI companies like OpenAI and Meta. Anthropic expressed satisfaction with the fair use ruling but did not comment on the piracy allegations.
5. <https://www.axios.com/2024/11/05/ai-training-data-publishers-ziff-davis> - A study conducted by Ziff Davis reveals that major AI companies, including OpenAI, Google, and Meta, heavily depend on content from premium publishers to train their large language models (LLMs), more than they publicly acknowledge. This reliance gives publishers leverage to demand copyright protection and compensation for their contributions to AI training. Executives at Ziff Davis, a member of the News/Media Alliance, have analyzed public datasets to show this disproportionate use of commercial publisher content. The findings are consistent with a previous report submitted by NMA to the U.S. Copyright Office. Unlike other major publishers who have secured data licensing deals with AI firms, Ziff Davis has not yet made such agreements. The study also notes that older public datasets are still being used in training newer LLMs. As deal-making trends shift, news companies are now focusing less on data training agreements, viewing them as one-time financial opportunities.
6. <https://www.asahi.com/ajw/articles/15822631> - In response to the recent rise of Retrieval-Augmented Generation (RAG) systems and other AI technology, many news organizations in Japan have set their websites to explicitly prohibit the use of their content by generative AI. Nevertheless, the association states there have been confirmed instances of AI firms ignoring these restrictions and unlawfully using association members’ content—including as clear references in RAG systems. Citing the Cultural Affairs Agency’s 2024 guidance on AI and copyright, the association noted that such unauthorized use likely constitutes copyright infringement. RAG systems, which use generative AI to search the web for information and generate responses to user questions, have seen a massive boom in popularity. However, this practice has raised concerns about a growing trend known as 'zero-click search,' where users read only the AI-generated responses without clicking through to the original news sources of the information. Media groups warn that, in addition to potentially generating misinformation, this can throttle traffic to legitimate news sites and ultimately threaten the sustainability of professional journalism.
7. <https://arstechnica.com/tech-policy/2025/05/copyright-office-head-fired-after-reporting-ai-training-isnt-always-fair-use/> - The head of the U.S. Copyright Office was dismissed after reporting that AI training isn't always considered fair use. The report highlighted that using pirated or illegally accessed content for AI training could harm the market for access to those works. The Copyright Office suggested that training on pirated collections of copyrighted works to build a training library, or the distribution of such a library to the public, would harm the market for access to those works. The report also noted that the use of pirated or illegally accessed material goes beyond established fair use boundaries. This development underscores the ongoing debate over the legality of using copyrighted content for AI training and the need for clearer guidelines.