



By: Tomorrow's Affairs Staff

# What impact will the BBC lawsuit have on the future of AI models?



The BBC sent a letter to the director of Perplexity AI, Aravind Srinivas, **demanding** an immediate end to content scraping (collecting and using material) from the public broadcaster's official website to train its model.

In addition, Perplexity must delete all previously downloaded content and submit a proposal for compensation for alleged copyright infringement. Research by the BBC's legal service found that Perplexity used the intact text of the original articles in 17 per cent of the responses generated, undermining public confidence in independent reporting.

Perplexity has **denied** these allegations in a statement to the Financial Times. The company claims that its models are not trained directly on publishers' content, but that the publicly available data collected is only aggregated during the processing stage.

In its response, Perplexity argues that the BBC has misunderstood the principles of web search and is applying intellectual property standards unfairly.

Previous cases prove that this is not the first time Perplexity has faced allegations of unauthorised content copying.

In June 2024, Forbes criticised the company for publishing excerpts from its articles without citing the source. In October of the same year, the New York Times sent a formal warning and demanded the suspension of the use of its articles for artificial intelligence training.

The WSJ and the New York Post **filed** lawsuits for copyright and trademark infringement. Perplexity subsequently agreed to revenue-sharing programmes, but negotiations did not result in a final settlement.

## A divide deepens

The dispute takes on added significance as the BBC negotiates with the UK government over the renewal of its public funding licence. A

public organisation must demonstrate that it protects its own sources and standards. A failure to prevent the unauthorised use of content could put its budget and reputation at risk.

The economic consequences of the dispute will also affect the startup community. Investors will be closely watching the outcome of the dispute, which could set a precedent for the collection of web content.

If the court upholds the BBC's claims, AI companies could be forced to purchase expensive licences to access the data. If not, lawmakers are expected to develop new regulations to match the rapid advancement of technology.

**Supporting the BBC's claims could trigger a wave of lawsuits from other media outlets**

More broadly, the conflict between the BBC and Perplexity reflects the divide between traditional media and technological innovators.

While publishers insist on strict copyright protection, startups often believe that the internet is a public good and that the free collection of publicly available data falls under fair use. This process could lay the groundwork for universal licensing standards and revenue-sharing programmes.

Supporting the BBC's claims could trigger a wave of lawsuits from other media outlets. Long court cases would force young engineers to prove the legality of their methods. On the other hand, a failure to protect rights would force publishers to demand more precise rules on the transparency of data collection.

Although the BBC has not officially announced its intention to build industry consensus through the dispute, the Public Service Board of Governors is considering tightening copyright rules while it negotiates with the government over the renewal of the public funding licence.

In July 2024, Perplexity **launched** the Publishers' Program, through which it shares a portion of advertising revenue with partners like Time Magazine and Der Spiegel. However, these contracts remain voluntary and are based on individual agreements, not statutory licences.

## Transparency vs. innovation

The UK's parliamentary culture and technology committees have announced that they are considering a new round of public consultations on the collection of web content for training AI models in 2025. This could include the precise definition of the boundaries of fair use and the obligation to maintain a register of data sources.

The European Commission adopted a law on artificial intelligence in December 2024 and is now developing technical **guidelines** for implementation, including rules on the transparency of data collection from the internet.



*The legal framework in the UK and EU is expected to set requirements for transparency in the data collection process in the next year - European Parliament*

In the United States, the US Copyright Office is **examining** the collection of publicly available web content for training models, while congressional committees are considering possible changes to copyright law.

Most AI companies use web scrapers - automated programmes that visit target web pages at regular intervals and remove HTML

tags. The resulting text is then broken down into smaller blocks and converted into numerical vectors that the models use to learn.

This way, AI startups can process billions of words per day and use them to optimise responses to specific queries. At the same time, however, this poses a challenge in terms of metadata accuracy and tracking content versions.

The legal framework in the **UK** and **EU** is expected to set requirements for transparency in the data collection process in the next year. The imposition of a detailed register of sources and versions of content used to train the model is a possibility.

This would require companies to implement complex systems to track changes on the web, further increasing operational costs and slowing down the pace of innovation.

At the same time, regulatory pressure could encourage the emergence of industry standards for revenue sharing. In the next six months, the largest publishers are expected to form a consortium to negotiate with leading AI companies. Their aim will be to establish a unique compensation model where revenue from subscriptions or advertising is shared in proportion to the quantity and quality of content downloaded.

Finally, it should be observed how the risk of litigation affects investors. If the courts set precedents that restrict the freedom to collect publicly available data, analysts estimate that the value of AI startups could fall by ten to thirty per cent over the next twelve months. On the other hand, the issue of responsible data collection and the payment of fees could strengthen user confidence in AI-based products in the long term.