# Google’s AI content usage policies force publishers into difficult trade-offs



Internal documents have emerged that illuminate Google's approach to content ownership and usage in the realm of artificial intelligence, revealing practices that have raised alarms about the company's growing dominance in both search and AI technologies. As highlighted in a recent report, Google appears to have limited publishers' options to prevent their content from being used in AI-generated summaries. This situation has sparked renewed scrutiny amid an ongoing U.S. antitrust trial.

The memos, attributed to Google Search executive Chetna Bindra, suggest a strategic choice by the company to prioritise monetisation of its AI services over respecting the autonomy of content creators. Publishers faced a stark ultimatum: they could either allow their content to contribute to AI products or opt out entirely, a move that would render their websites invisible in Google's vast search ecosystem. This tactic effectively undermines the ability of publishers to control their digital footprint while emphasizing Google's aim to capitalise on AI-generated summaries as a lucrative venture.

Although Google launched a policy in 2023 ostensibly allowing publishers to opt out of content usage for AI training, this policy has glaring loopholes. The recent court hearings revealed that it did not extend to significant features like AI Overviews embedded in Google Search. The only genuine way for publishers to block their content from being scraped for these features is to ban Google's crawlers altogether—an action that would eliminate their visibility in search results. This stark reality presents a significant dilemma for publishers reliant on Google for traffic, highlighting the tension between participation in the digital ecosystem and the control over their intellectual property.

Further complicating this landscape, it has been reported that multiple AI companies are flouting the widely accepted Robots Exclusion Protocol. This protocol is designed to enable publishers to dictate which parts of their site can be accessed by web crawlers. However, companies are bypassing these signals, raising serious concerns about the potential erosion of content monetisation and the future viability of journalism itself. Legal disputes are emerging as major publishers, including The New York Times, seek redress for unlicensed use of their content in AI training.

The backlash against Google and other tech giants is escalating, as a growing number of websites are actively blocking AI crawlers from accessing their content. According to recent data, en masse, nearly 20% of the world's top websites have enacted measures to keep AI bots at bay. This shift represents a reaction to the ambiguous legal landscape surrounding the use of copyrighted materials by AI systems and reflects a broader wariness among publishers toward AI’s encroachment on their business models. While historically, many publishers welcomed search engine traffic, increasing doubts regarding AI’s implications are prompting a reassessment of this relationship.

In response to these mounting concerns, Google recently tested the removal of links to California news websites from search results in retaliation to proposed legislation aimed at requiring tech companies to pay for the use of media content. This move highlights a strategy by Google to assert its influence in legislative matters, further complicating the relationship between digital platforms and traditional news publishers. Critics argue that such actions from Google may threaten the sustainability of local journalism, as content creators struggle to navigate a landscape where their work is extracted without consent.

Moreover, Google’s woes with regulators are not limited to the U.S. The French competition watchdog charged the company with a hefty fine for breaching EU intellectual property regulations, particularly regarding its AI chatbot. Google agreed to a settlement acknowledging it had improperly used content from publishers, underscoring the increasing scrutiny of its practices by regulatory bodies internationally. Publishers are not just voicing misguided grievances; they are taking a stand against a system they perceive as exploitative.

As the industry grapples with these dynamics, many fear that the balance between innovation and ethical content usage is tipping unfavourably. Google's transformation towards AI-driven results is poised to reshape the internet fundamentally, with implications that could reverberate across all sectors reliant on digital presence. Careful management of this transition is essential to prevent the erosion of the very content and contributors that have historically enriched the digital landscape.

## Reference Map:

* Paragraph 1 – [[1]](https://www.timesnownews.com/technology-science/google-made-opting-out-of-ai-search-nearly-impossible-for-publishers-article-151708644), [[6]](https://www.axios.com/2024/05/17/google-openai-ai-generative-publishers)
* Paragraph 2 – [[1]](https://www.timesnownews.com/technology-science/google-made-opting-out-of-ai-search-nearly-impossible-for-publishers-article-151708644), [[2]](https://www.reuters.com/technology/artificial-intelligence/multiple-ai-companies-bypassing-web-standard-to-scrape-publisher-sites-licensing-2024-06-21/), [[4]](https://apnews.com/article/0921cc2b39de591c173201e1ec4bee64)
* Paragraph 3 – [[5]](https://www.reuters.com/technology/french-competition-watchdog-hits-google-with-250-mln-euro-fine-2024-03-20/), [[3]](https://www.axios.com/2023/08/31/major-websites-are-blocking-ai-crawlers-from-accessing-their-content)
* Paragraph 4 – [[1]](https://www.timesnownews.com/technology-science/google-made-opting-out-of-ai-search-nearly-impossible-for-publishers-article-151708644), [[3]](https://www.axios.com/2023/08/31/major-websites-are-blocking-ai-crawlers-from-accessing-their-content)
* Paragraph 5 – [[6]](https://www.axios.com/2024/05/17/google-openai-ai-generative-publishers), [[5]](https://www.reuters.com/technology/french-competition-watchdog-hits-google-with-250-mln-euro-fine-2024-03-20/)
* Paragraph 6 – [[5]](https://www.reuters.com/technology/french-competition-watchdog-hits-google-with-250-mln-euro-fine-2024-03-20/), [[4]](https://apnews.com/article/0921cc2b39de591c173201e1ec4bee64)
* Paragraph 7 – [[3]](https://www.axios.com/2023/08/31/major-websites-are-blocking-ai-crawlers-from-accessing-their-content), [[2]](https://www.reuters.com/technology/artificial-intelligence/multiple-ai-companies-bypassing-web-standard-to-scrape-publisher-sites-licensing-2024-06-21/)

Source: [Noah Wire Services](https://www.noahwire.com)

## Bibliography

1. <https://www.timesnownews.com/technology-science/google-made-opting-out-of-ai-search-nearly-impossible-for-publishers-article-151708644> - Please view link - unable to able to access data
2. <https://www.reuters.com/technology/artificial-intelligence/multiple-ai-companies-bypassing-web-standard-to-scrape-publisher-sites-licensing-2024-06-21/> - Multiple AI companies are bypassing the Robots Exclusion Protocol (robots.txt) to scrape content from publisher sites, as highlighted by content licensing startup TollBit. This protocol is widely used by publishers to control which parts of their website can be accessed by web crawlers, but some AI companies are ignoring it to extract content for their generative AI systems without permission. A case in point is the dispute between AI search startup Perplexity and Forbes. The News Media Alliance, representing over 2,200 U.S. publishers, expressed concerns that ignoring these "do not crawl" signals could undermine content monetization and the journalism industry. TollBit, which matches AI companies with publishers for content licensing, observed that multiple AI agents are bypassing the protocol, looking to monetize valuable content like news and exclusive insights. Legal actions and licensing agreements are being pursued by some publishers, including the New York Times, against AI companies for content use without consent. The debate over compliance and content value continues amid growing use of AI-generated summaries, exemplified by Google's AI summary product affecting search visibility.
3. <https://www.axios.com/2023/08/31/major-websites-are-blocking-ai-crawlers-from-accessing-their-content> - According to data from Originality.AI, nearly 20% of the world's top 1000 websites are preventing AI crawlers from accessing their content. This trend is rising as websites react to the lack of clear legal guidelines on AI's use of copyrighted material. For example, OpenAI introduced its GPTBot crawler in early August, instructing websites on how to block it. The number of sites blocking OpenAI's ChatGPT bot grew from 9.1% on Aug 22 to 12% by Aug 29. Crawler bots, which operate like web browsers but store data instead of displaying it, are controversial. While companies like Google argue their use constitutes fair use, many publishers disagree and have taken legal action. Some publishers had initially allowed search engines to crawl their sites due to the traffic they brought in, but now view AI services with skepticism. This growing blockade against AI crawlers could complicate efforts to refine AI products, as good data becomes harder to obtain. News companies are particularly torn between embracing and resisting AI.
4. <https://apnews.com/article/0921cc2b39de591c173201e1ec4bee64> - Google has started removing links to California news websites from search results for some users, in response to a proposed California law that would require tech companies to pay media firms for linking to their content. This test serves as a warning to the state Legislature, which is considering a bill mandating compensation to media companies from tech giants like Google, Facebook, and Microsoft. The bill aims to mitigate the decline in journalism jobs by providing a revenue stream to struggling legacy media companies. Google claims the legislation disrupts its existing model and threatens to pause investments in the California news industry. Though supporters say the bill would provide financial support to local news organizations, opponents argue it could harm local journalism and benefit out-of-state newspaper chains. The outcome of this legislative battle could significantly impact the relationship between news publishers and large digital platforms.
5. <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.
6. <https://www.axios.com/2024/05/17/google-openai-ai-generative-publishers> - La transformación de Google hacia resultados generados por IA, reemplazando la familiar lista de enlaces, está reconfigurando internet y podría acelerar el declive de la World Wide Web de más de 30 años. Desde la aparición de ChatGPT a fines de 2022, Google ha aumentado constantemente sus resultados generados por IA y esta semana anunció la implementación de "Resumenes de IA". Los editores y minoristas temen que esto afecte profundamente su tráfico de referencia y, potencialmente, destruya sus negocios. En el evento I/O de Google, se mostró la capacidad de Google Lens para interpretar imágenes de videos en vivo, demostrando una mayor eficiencia en la búsqueda de respuestas en comparación con la última década. Sin embargo, la web actual existe gracias a la contribución de millones de personas que han añadido conocimiento, tradición e imágenes con diversos objetivos. Aunque las plataformas de redes sociales comenzaron el desmantelamiento lento de la web abierta, la introducción de resúmenes de IA de Google podría acelerar su declive. Es esencial que Google maneje este cambio con cuidado para evitar que la IA consuma su propia fuente de sustento.