



Strategic Market Status Investigation into Google’s General Search and Search Advertising services

Comments from the European Publishers Council (EPC) 3rd February 2025

The EPC[1] welcomes the CMA’s invitation to comments (“ITC”) on the scope of, and issues to be explored in, its SMS investigation of Google’s General Search and Search Advertising and Search Advertising Services.

The EPC supports the approach that the CMA proposes to pursue in its investigation and considers that the ITC lays out all the key issues that the CMA should investigate when designating Google Search as having strategic markets status (“SMS”).

This submission is divided into three parts. First, we explain the key issues faced by press publishers with respect to Google’s significant market power in search and search advertising (**Section I**). Then, we comment on the scope of the investigation and SMS assessment (**Section II**). Finally, we comment on the issues the CMA intends to explore and the possible interventions it may pursue in search (**Section III**).

I. Issues faced by press publishers with respect to Google Search

Google and press publishers are in a relationship of interdependence. On the one hand, Google needs access to press content as such content draws users to Search (and other related services), which Google can then monetise in various direct, and indirect ways. On the other hand, press publishers benefit from referrals from Search, in that Search users may click on the links to publishers’ content that are provided by Search.

While Google and press publishers benefit from their interactions, Google’s market power has allowed it to draw more value from these interactions to the disadvantage of press publishers. **Search has progressively transformed from a referral engine to an engine**

directly responding to the user queries, through a series of features such as the “Knowledge Panel”, “Featured Snippets”, “People Also Ask”, and more recently “AI Overviews”, which was launched in the UK in April 2024. These features aim to keep users on the Search Engine Results Page (“SERP”), hence increasingly depriving press publishers from referrals as well as advertising revenues generated by direct audiences.

Furthermore, it has been well documented that **Google’s own services continue to receive better placement than Google’s business users**; Google’s self-preferencing conduct not only harms intermediaries in the fields of travel, shopping, and transport. It further harms content service providers, including news publishers, which rely on Google’s search engine to reach their audiences. Google’s discriminatory behaviour has significant effects on traffic, ad revenues, and clicks. These effects are felt by all third-party providers that compete with Google for users’ attention. However, where news publishers are on the receiving end of Google’s practices, media pluralism and the citizens’ right to receive information are also undermined.

AI Overviews is a particularly worrying product for publishers. Google’s AI Overviews are derived from a combination of the language generation capacity and contextual understanding of their LLM, and **the real-time capture of web data as follows:**

1. LLM Technology:

- The core of AI Overviews relies on large language models (LLMs) trained on vast datasets, which include publishers’ content which was ingested without permission or compensation to generate natural language responses.
- These models can synthesise information, identifying patterns, and generating coherent summaries based on their training.

2. Real-Time Web Data Retrieval:

- In addition to the LLM’s pre-trained knowledge, **AI Overviews pull in real-time information from the web, through a combination of Google's dominant search indexing and live data retrieval power.** This ensures that the summaries are current and relevant, especially for topics that require up-to-date information like news, events, or trending topics which then compete head on with publishers’ own sources.

Furthermore, **these AI-generated answers come on top of the search results**, pushing down any links to the original sources on the SERP and **there is no option for users to opt-out.**

Thus, instead of being presented with a selection of links and brief snippets in answer to a search query, consumers are presented with an AI-enabled summary drawn from multiple sources to produce a curated response, eliminating the need for users to click-through to original content and, therefore, syphoning off monetisation opportunities, and keeping users in the environment of the gatekeeper[2] thereby preventing publishers from building long term audiences.

These “answers” are a massive leap beyond the way search currently works, and threaten to further entrench the position of large platforms as information gatekeepers, while, in addition, **creating legal ambiguity as to their role as information creators/producers for which they benefit from the ‘platform safe-harbours’** i.e. limitations of legal liability for that AI produced content unlike publishers with whom they compete unfairly. Furthermore, AI Overviews shift Google’s role beyond simply selling ads on traditional search results that direct users to the open web. Instead, they monetise highly targeted ads within AI-generated summaries, which reduce visibility and direct access to publishers on the open web. Tech research firm Gartner predicts traffic to the open web including publishers’ websites from search engines will fall by 25 percent by 2026.[3]

Furthermore, **when a dominant firm such as Google integrates AI Overviews trained on copyrighted material without consent this should not be viewed purely as a copyright issue;** it is a competition issue with broad market implications. By embedding unlawfully trained AI into essential digital products like browsers, dominant firms such as Google:

1. Gain an unfair advantage over competitors.
2. Harm independent creators and publishers by devaluing their work.
3. Foreclose competition in emerging AI markets.

Competition authorities should factor copyright infringement into antitrust enforcement to prevent dominant firms from leveraging AI to entrench their power. Allowing this behaviour unchallenged would set a dangerous precedent where breaking the law becomes a competitive strategy.

In theory, press publishers can block Google’s web-crawler, Googlebot, which scrapes content for AI purposes, including for AI Overviews. However, since Google uses the same crawler for both AI Overviews as it does for traditional search, doing so would be self-defeating as the blocking of the crawler would also have an adverse impact on the referencing of publishers’ content on Google search, with the consequences of Google’s

referred traffic going to other websites without having any disciplinary effect on Google’s conduct. Furthermore, crawler blocking is unsuitable as a solution for managing copyright in AI training, as it requires publishers to block AI crawlers (e.g., through robots.txt) which places the responsibility on them, rather than on AI developers such as Google who should be obtaining proper licenses before using copyrighted content in accordance with publishers’ rights reservations. **Specifically, to the issue of crawlers, EPC has developed proposals for responsible crawling (enclosed).** The development of GenAI services and the integration of these by gatekeepers is explored more fully in the EPC’s Vision Paper pages 7-11 available at <https://www.epceurope.eu/post/vision-paper-2025>

Due to Google’s dominant market power, **there is a clear imbalance in the relationship between Google and press publishers as recognised by regulators.** For instance, OFCOM and the CMA observed in their Joint advice to DCMS on Platforms and Content Providers (“OFCOM/CMA Joint Advice”): *“There is an imbalance in bargaining power between large platforms and content providers, including news publishers. Each publisher needs Google and Facebook more than the platforms need them.”* And “[g]iven that we estimate that there are hundreds of online UK brands with news content, it is likely that individual news publishers will be to some extent substitutable from the platform perspective.” [4] Thus, something needs to be done to address this imbalance, and the DMCC presents the right opportunity to do so.

To appease news publishers, Google launched Google News Showcase in October 2020 which is a thinly veiled attempt to mitigate regulatory scrutiny. News Showcase is not a genuine user-facing product, but rather a contracting tactic as exposed by the French Competition authority (*Autorité de la concurrence*). In its decision on March 15, 2024, the Authority imposed a €250 million fine on Google for failing to comply with previous commitments to negotiate in good faith with press publishers. One of the breaches identified was related to Google’s AI service, **Bard** (later renamed **Gemini**). The Authority found that Google used content from press publishers and news agencies to train and operate Bard without informing or negotiating with the rights holders, which was deemed a failure to negotiate in good faith. Despite being created by Google, Showcase has very low user engagement and delivers negligible traffic to the press publishers. The [REDACTED] hence making it urgent that the CMA sets up a mechanism that will allow press publishers to negotiate with a view to receiving fair compensation for Google’s use of their content.

II. Scope of the investigation and SMS assessment

We support the scope of the investigation and the fact that the CMA will look both at Google's search and search advertising businesses and treat them as a single activity. It is obvious that they go together and insufficient competition in search has an adverse impact on search advertising.

More specifically, we agree that:

"Google's general search activity includes all the user-facing aspects of its search engine: allowing users to search the world wide web through any medium (including but not limited to websites, web browsers, smart speakers, and **AI interfaces**); and returning results in the form of information of any type (including but not limited to all the information on Google's SERP: organic and paid-for search results, the Google Discover feed, and search features such as links to specialised search services, videos and maps, and **AI Overviews**)."[5] (emphasis added)

It is critical in our view that the CMA includes products like AI Overviews its analysis, and we are very pleased it intends to do so, given that AI Overviews are now prominently displayed on the Google Search results page and as noted above, without the possibility for users to opt out.

"Google's search advertising activity includes all the business-facing functionality and services supporting search advertising, including Google Ads and SA360."[6]

We hope and expect that the CMA will designate Google as having SMS in its Search activity at the end of this investigation. We consider that the findings made in the CMA's 2020 Market Study remain largely relevant. Google Search continues to enjoy significant economies of scale and network effects and has access to vastly more data than any other competitor. As illustrated by the recent judgment of the US District Court for the District of Columbia in *US v. Google LLC*, [7] Google has monopoly power in general search services, and has engaged in a variety of tactics to maintain and expand its market power in general search services with significant implications in the search advertising market.

It is likely that, as in the abovementioned US case, Google will argue that its market power has weakened given the development of generative AI. But the US District Court based on the extensive evidence presented to it, rejected this argument observing that:

“the advent of artificial intelligence (AI) has not sufficiently eroded barriers to entry—at least not yet. New technologies may lower, or even demolish, barriers to entry, but such innovation is meaningful only if it can change the market dynamic in the ‘foreseeable future.’ ... Currently, AI cannot replace the fundamental building blocks of search, including web crawling, indexing, and ranking.”[8]

We certainly agree with this view. There is indeed no evidence that Google’s market power in search and search advertising has been dented by the AI-enabled services (chatbots, etc.) developed by rivals. In addition, **Google is a major developer of AI itself, and as described above has supplemented and enhanced Search with Generative AI.** There is, therefore, no case for the proposition that Google should escape designation on the basis that Generative AI may potentially reduce its market power.

The EPC agrees with the avenues of investigation set out in paragraphs 26-28.

III. Issues explored and possible interventions

Google’s significant market power in search and search advertising, and the ways in which Google has leveraged market power has been a source of concerns for thousands of UK businesses and consumers, and the subject of multiple investigations and lawsuits across Europe, which for reasons largely expressed in the Furman report[9] have not provided robust responses to these concerns. As the above concerns have been unaddressed for so long, the EPC encourages the CMA to impose the first set of conduct requirements at the same time as its designation decision in October 2025 or soon thereafter.

As to the issues to be explored, the EPC agrees with:

- The need to **address weak competition and barriers to entry and innovation in search.** Most concerns relating to search and search advertising result from the lack of competition in these activities, and both businesses and end users will benefit from greater contestability in search.[10]

- The need to **prevent leveraging market power and ensuring open markets**. In particular, the EPC is pleased that the CMA will investigate the restrictions on the tying of Google’s search web crawling with web crawling for the purposes of its AI services. We agree that given Google’s strong position in general search, firms have *“no realistic option but to permit access to the web crawlers that populate Google’s search results, in order to ensure that they appear in search results and are therefore visible to their customers.”* This does not only give Google *“an advantage over competing AI developers”*,^[11] but also allows Google to exploit press publishers by scraping their content for the development of its AI services, which, as explained in Section 1 above with respect to AI Overviews, will deprive the ability of press publishers to monetize their content. Thus, Google’s tying conduct is not only a source of exclusion, but also one of exploitation.

- The need to **protect users against exploitative conduct**. In particular:
 - a) We support the imposition of requirements on Google to **give consumers more control over their data**.^[12]

 - b) The CMA should impose requirements on Google to **ensure search rankings are non-discriminatory**. This is particularly important for press publishers which play an essential role in the democratic process. It is therefore particularly important that Google does not discriminate against certain publishers (on the ground that they have expressed their discontent against its policies) or certain views.^[13] It is also why we support the view that there is a need for an effective complaint mechanism in case a press publisher (or any other business whose content appears in Search) considers it has been subject to discrimination in rankings.^[14] We emphasise that such a mechanism should not be perfunctory.

 - c) Most importantly, the EPC strongly supports the imposition of **requirements to ensure fair terms (including payment terms) for use of publisher content**.^[15] The CMA should rely on the findings of its joint advice with OFCOM, which are compelling. There is considerable evidence that press publishers do not receive fair terms from Google when Google uses their content. They are unable to capture their fair share of the “joint value” that is created by Google’s use of their content for Search and other news related services (Google News, Discover, etc.). This is particularly distressing at a time where press publishers are under a great deal of financial pressure.

- d) We also agree that **these requirements could (or in fact should) include both payment and non-payment terms, and we strongly encourage the CMA to “explore issues that have been raised more recently relating to the use of publisher content in Google’s AI Overviews and in grounding responses of AI assistants.”**[16] AI Overviews represents a major threat for publishers, which are unable to obtain fair terms from Google, including for the licensing of their content used by this service. As noted above, their valuable content is used without attribution and without any possibility to monetise it. Because AI Overviews presents summaries of news content, there is a considerable risk that many users will dispense from consulting original sources, with the mid to long-term resulting in publishers no longer being able to build their audiences. Furthermore, because the AI generated 'answers' are in effect personalised, they are difficult to monitor meaning Google runs limited risk as to the answers it shows (even if they contain obvious mistakes that could damage a party), whereas publishers are held accountable for information published. The lack of accountability of the AI provider creates an uneven playing field.
- e) As far as payment terms are concerned, the EPC would like to emphasise **the necessity for conduct requirements to include a backstop mechanism that applies when Google and news publishers are unable to agree on terms.** The availability of a **final offer arbitration mechanism** in such situations may, for instance, prevent Google from delaying negotiations or pursuing them in bad faith in order to create a stalemate.
- f) We note the CMA consideration of requirements for Google **to make available key data to competitors.** EPC would be concerned that such measures risk compounding and perpetuating the harm to publishers by enabling third parties to benefit from the scraping and indexing of publishers’ content on unfair terms.

Conclusion

The EPC strongly supports the DMCC regime and is impressed with the careful thought that has gone into the ITC. We look forward to engaging on these issues which are of vital importance for the press publishing industry.



-
- [1] The European Publishers Council (EPC) is a high-level group of Chairmen and CEOs of Europe's leading media groups representing companies which are active in news media, television, radio, digital market places, journals, eLearning, databases and books.
- [2] EPC 2030 Vision Paper, Digital Evolution in Media: Safeguarding Democracy by Investing in Quality Journalism for a Sustainable Future, available at file:///C:/Users/Damien/Downloads/EPC%202030%20Vision%20Paper_final.pdf
- [3] See Will traffic from search engines fall 25% by 2026?, available at <https://searchengineland.com/search-engine-traffic-2026-prediction-437650>
- [4] Platforms and content providers, including news publishers Advice to DCMS on the application of a code of conduct, November 2021, available at https://assets.publishing.service.gov.uk/media/6273af6be90e0746c882c361/Platforms_publishers_advice_A.pdf
- [5] ITC, para. 21.
- [6] Id., para. 22.
- [7] See <https://oag.dc.gov/sites/default/files/2024-11/2024.11.20%20Google%20Plaintiffs%20Initial%20PFJ%20%281%29.pdf>
- [8] Id. at page 167.
- [9] [Unlocking digital competition, Report of the Digital Competition Expert Panel, 2019, https://www.gov.uk/government/publications/unlocking-digital-competition-report-of-the-digital-competition-expert-panel](https://www.gov.uk/government/publications/unlocking-digital-competition-report-of-the-digital-competition-expert-panel)

- [10] ITC, para. 39.
- [11] Id, para. 42(b).
- [12] Id., para 43(a).
- [13] Id., para 43(c).
- [14] Id., para 43(b).
- [15] Id., para 43(d).
- [16] Id.

Members of the European Publishers Council have been working with our experts to develop a proposal for Mandatory Standards for Responsible Crawling set out below:

Proposed standard for responsible crawlers in AI development

27 November 2024

Introduction

As AI continues to evolve, the use of **automated web crawlers** by AI companies to gather data has increased significantly. However, this unregulated practice raises substantial ethical and legal concerns, particularly around intellectual property, privacy, and the transparency of AI systems. Consequently, there is a growing need for a codified set of mandatory standards that define how AI companies should develop and deploy crawlers in full compliance with any rights reservations on websites to be crawled.

These standards could form part of the transparency obligations under the AI Act and be included in the forthcoming Code of Practice as part of the Copyright Policy for each AI company. This will ensure compliance with established rights and ethical norms while not stifling innovation in AI. The establishment of a quality trust mark could signify adherence to these standards, helping to build trust in AI systems.

Non-compliance with these standards should be treated as copyright infringement, thus creating a robust framework to protect intellectual property and giving content creators a legal tool to safeguard their work. The core elements of these standards would centre on the transparency and accountability of AI crawlers and respect for intellectual property. They will include clear obligations for identity disclosure, purpose declaration and an audit trail of activities. These standards will help to promote responsible AI development.

Transparency and accountability framework

The foundation of these mandatory standards would be a framework emphasising transparency and accountability in line with the AI Act's transparency obligations. Each AI crawler must be identifiable, transparent about purposes for crawling, auditable, and respectful of webmaster's choices regarding access to their content. This framework could include the following headings:

1. **Traceability** - Every crawler must have a distinct, consistent "user-agent string" and a transparent identity traceable back to the AI company or third party. Crawlers should be prohibited from using another crawler's name to present themselves as legitimate bots pertaining to well-known companies.

2. **Purpose** - Each crawler must link to a statement (“a Declaration”) that details its purpose, provides specific information on how the collected content will be used gives storage duration (as this impacts the long-term implications of data use). This Declaration should cover aspects such as whether the data will be used for indexing, training machine learning models, or other purposes.
3. **Separation of roles** - Crawlers for differing functions, such as indexing and training, must be separate to avoid crossover of collected content for unintended uses - so for example, content for indexing is not inadvertently used in model training. By offering explicit statements on the crawler’s purpose, AI companies can help alleviate concerns around unauthorised data use and demonstrate responsible practices.
4. **Accountability** - A Data Hub where AI companies deposit information on each of their crawlers' identities and purposes should be created. The Hub would be publicly accessible and include each crawler's information, including the unique "user-agent string" and identity of its owner. This registration process ensures that AI crawlers are easily identifiable by content creators and website owners, fostering transparency across the industry. Any third-party crawlers used by AI developers should also be added to the Data Hub, with the responsibility for compliance resting on the AI developer deploying them. If any changes to the crawler’s identity are made, the information provided to the Hub must be updated before the crawler is used. This transparent approach strengthens trust in AI and ensures that all entities collecting data are clearly accountable. The Data Hub could be managed by the EU Commission or the EUIPO.
5. **Audit trail** - A mandatory audit trail will enable accountability by providing a record of each crawler's activities. Companies should maintain a log of all content accessed, including timestamps, and any subscription-based or content behind paywalls, so that any claims of unauthorised access can be investigated and addressed. Additionally, this audit trail aligns with transparency requirements established by the AI Act, as it allows regulators and stakeholders to assess compliance and ensure the integrity of AI systems.

Obligations for website terms and conditions

AI companies have a responsibility to respect website-specific terms and conditions including machine-readable rights reservations as well as exclusion and access protocols such as "robots.txt" files. By respecting these standards, AI developers demonstrate their commitment to responsible data practices and uphold the rights of content creators and publishers. This obligation should go hand in hand with a commitment to work on a new ai-specific protocol to update robots.txt: a standardised protocol that would allow webmasters

to define AI-specific permissions and exclusions. This would help streamline compliance and provide a straightforward, consistent framework for website owners to manage AI crawlers' access to their content.

Enforcing compliance through copyright and intellectual property protection

To ensure adherence to these standards, non-compliance should be treated as a copyright infringement, providing content creators with a legal recourse for unauthorised use of their work. This approach will create strong incentives for AI companies to respect rights reservation signals and follow best practices for ethical data gathering. Additionally, it emphasises the importance of consent in using online content, acknowledging the intellectual property rights of website owners and content creators.

The role of a quality trust mark

A quality trust mark could serve as a symbol of compliance and accountability, signalling that an AI company adheres to the responsible crawling standards as part of the Code of Practice. This mark would provide a competitive advantage for companies committed to ethical data use, offering a level of trust to consumers, partners, and regulators. It would help ensure the AI developers are complying with the requirements – with compliance assessments either undertaken by the EU AI Office or outsourced to a third party.