

CMA proposed conduct requirements in respect of Google's general search services

Response from DMG Media

1. DMG Media is the publisher of the Daily Mail, Mail on Sunday, Metro and i newspapers; MailOnline, metro.co.uk and inews.co.uk websites; and New Scientist magazine.
2. DMG Media welcomes the opportunity to respond to the CMA's consultation on the first batch of proposed conduct requirements for Google Search. We continue to support the CMA in its careful and thorough work in this area. We recognise the CMA is operating in a complex and highly contested political environment. However, we consider targeted, proportionate amendments to the CRs could materially enhance their workability.
3. In respect of the Publisher Conduct Requirement (CR),¹ in order to make the CR effective, the CMA should:
 - a. **Strengthen transparency**, requiring publisher-specific reporting on purpose-level and product-level use, zero-impression grounding and URL-level data access. Publishers should also have access to user engagement data on a per-product basis, for example click-through rate and impressions data separately as between AI Overviews and AI Mode.
 - b. **Clearly distinguish grounding, fine-tuning and training for generative AI services and features**,² and allow publishers to opt in to each separately without adversely impacting their ranking in general search. These are distinct use cases and treating them separately is necessary to prevent circumvention through (for example) continuous fine-tuning.
 - c. **Enable opt-in on a per-product basis**, preserving publisher control and preventing future bundling of Google's products.
 - d. **Close the overseas dataset loophole**, by clarifying that where Google "legally obtains" publisher data from open-source datasets, those datasets must have complied with UK copyright law in scraping the data.
 - e. **Pro-active opt-in should be required**, meaning Google should obtain publishers' express opt-in consent before their content can be used in AI services, similar to the position for publishers seeking user consent under the General Data Protection Regulation. At a minimum, opt-out controls should be simple, prominent and frictionless to exercise. They should be clearly signposted, proactively

¹ CMA, *Consultation: Publisher Conduct Requirement* (28 January 2026) available [here](#).

² CMA, *Consultation: Publisher Conduct Requirement* (28 January 2026), page 15, available [here](#); "**generative AI services and features** means Google's search generative AI features and broader generative AI services, together". Nb "... **broader generative AI services** means Google's generative AI-dependent products and services offering information retrieval capabilities outside of general search, such as Gemini AI Assistant and the Vertex AI API."

communicated by Google and accessible through straightforward, standardised mechanisms that do not require technical expertise or involve delay.

- f. **Consider separation of Google’s crawlers by purpose (search, grounding, fine-tuning, training)** to give publishers direct control and real-time data in respect of access to their content, facilitate content marketplace adoption and reduce reliance on CMA monitoring.
 - g. **Close the Search API monetisation gap**, by preventing Google from monetising access to publisher content via Search or grounding APIs third-parties to use in their own AI services. This also supports the facilitation of marketplaces, by enabling Google or third parties to channel buyer demand (which is the model adopted by Microsoft).
 - h. **Require independent auditing of Google’s six-monthly compliance reports.**
 - i. **Develop a payment-for-content CR now**, in parallel with the current package, rather than waiting 12 months.
4. As explained further in DMG Media’s submission, the recommendations at 3(a), (b), (c), (f) and (g) will strongly support the adoption and development of sustainable content marketplaces.
5. In respect of the Fair Ranking CR, the CMA should:³
- a. **Explicitly prohibit the Site Reputation Abuse Policy**, rather than relying on an ex-post complaints process that has no clear way to prevent harmful policies.
 - b. **Introduce explicit and binding minimum notice periods for material changes to ranking policies** (one month for routine changes; six months for major changes).
 - c. **Strengthen the complaints mechanism**, including clear timelines for responding to complaints, a requirement for reasoned decisions and wider access to independent appeal.

³ CMA, *Consultation: Fair Ranking Conduct Requirement* (28 January 2026), available [here](#).

Publisher CR

Publisher control – CR(2) and CR(3)

6. DMG Media supports the broad purpose of Publisher CR(2)⁴ and (3)⁵, which the CMA states is to provide “sufficient controls over Google’s use of publishers’ Search Content in its generative AI services and features.”⁶ For too long, Google has been able to leverage its market power in Search to force publishers to provide access to valuable content for free. DMG Media welcomes the CMA grappling with this issue. That said, DMG Media is concerned the CR is ineffective. In this respect it has seven significant recommendations, which it has set out below.

Recommendation 1 – The default position should be that publishers are opted out

7. Publishers should not be required to take active steps to prevent their content being used in Google’s AI services. Instead, publishers should be opted out by default. Google should be required to seek publishers’ express opt-in through, for example, a consent banner that appears when a publisher opens the Google Search Console. Publishers are under comparable obligations regarding user consent under the GDPR, and this approach would be consistent with those principles. Having opt-out as the default would also strengthen the bargaining power of publishers by reducing the risk that content is used by default or through inadvertent consent, and by ensuring use occurs only following an active decision to opt in.
8. At a minimum, if the CMA does not require opt-in, controls for opting out should be prominent, clearly signposted and proactively communicated by Google. Any opt-out controls should be easy to access through straightforward, standardised mechanisms and simple to exercise without technical complexity. The process should not involve undue friction or delays. This issue is not currently covered at all in the CR, which leaves room for circumvention.

⁴ CMA, *Consultation: Publisher Conduct Requirement* (28 January 2026), page 15, available [here](#); “2. Google shall provide publishers with effective controls to withhold their Search Content from being used in: a. the training and grounding of its broader generative AI services; and b. the grounding of its search generative AI features.”

⁵ CMA, *Consultation: Publisher Conduct Requirement* (28 January 2026), page 15, available [here](#); “3. In relation to the controls described in paragraph 2, Google: a. shall ensure that those controls evolve in an appropriate way as generative AI services and features evolve; b. shall not: i. maintain or introduce ranking signals whose purpose is to downrank opted-out Search Content in general search outside of search generative AI features; or ii. Cause Search Content to be presented or displayed differently in general search outside of search generative AI features depending upon whether it is opted or out through those controls; and c. shall not attempt to circumvent any publisher’s choice to withhold its Search Content by acquiring that content through other sources.”

⁶ CMA, *Consultation: Publisher Conduct Requirement* (28 January 2026), page 23, available [here](#).

Recommendation 2 – Publishers should be able to opt into training for search generative AI services and features

9. The CRs currently provide that Google must enable publisher control over use of content for: (a) training and grounding for **broader generative AI services**⁷; and (b) grounding, but not training, for **search generative AI features**.^{8,9}
10. DMG Media’s view is that publishers should be able to opt in to allow Google to use its content for training of search generative AI features (i.e. AI Overviews and AI Mode). Publishers’ content constitutes highly valuable input data that Google uses to train its search generative AI features and generate revenue. Publishers must have the right to withhold that value.
11. Any suggestion that a decision not to opt in may affect search ranking is not a legitimate objection.¹⁰ Preventing Google from leveraging ranking as a means of coercing access to publisher content was a central rationale for introducing the CR in the first place. Ranking cannot be used, directly or indirectly, as a rationale for not providing control.
12. Finally, we do not accept that extending opt-in to training for search generative AI features without opt-in affecting ranking is technically infeasible. Any advantage Google gains by bundling its fine-tuning crawler to its search rankings is an unfair competitive advantage not enjoyed by other AI companies. Additionally, no evidence has been provided in the CMA’s consultation document to substantiate this claim and we strongly suspect Google could isolate fine-tuning without impacting its ranking algorithms were it not commercially incentivised to refuse to do so.

Recommendation 3 – Publishers should be able to opt in to fine-tuning, as a distinct activity, for both search generative AI features and broader generative AI services

13. Publishers should be able to opt into fine-tuning separately, for both search generative AI features and broader generative AI services, and these activities should be clearly distinguished in the CRs from training and grounding. The current drafting is unclear on whether opting out of “fine-tuning” was intended to be covered in the requirement that publishers have control over use of their content in “training” for broader generative AI services, but, in any case, we consider this should be explicitly distinguished. We consider this is equally the case for search generative AI features.
14. Fine-tuning is a distinct stage that sits between general model training and grounding. It involves the use of content to optimise AI systems for particular objectives. Unlike grounding, which does not alter model weights and has no lasting effect beyond an individual response (unless the grounded response is fed back into the AI

⁷CMA, *Consultation: Publisher Conduct Requirement* (28 January 2026), page 15, available [here](#); “**broader generative AI services** means Google’s generative AI-dependent products and services offering information retrieval capabilities outside of general search, such as Gemini AI Assistant and the Vertex AI API”

⁸ CMA, *Consultation: Publisher Conduct Requirement* (28 January 2026), page 15, available [here](#); “**search generative AI features** means Google’s generative AI-dependent features offered within general search, such as AI Overviews and AI Mode.”

⁹ CMA, *Consultation: Publisher Conduct Requirement* (28 January 2026), page 15, available [here](#); “2. Google shall provide publishers with effective controls to withhold their Search Content from being used in: a. the training and grounding of its broader generative AI services; and b. the grounding of its search generative AI features.”

¹⁰ CMA, *Consultation: Publisher Conduct Requirement* (28 January 2026), page 25, available [here](#).

system for training or fine-tuning purposes), fine-tuning permanently modifies model parameters and embeds learnings derived from publisher content. Through repeated testing, user evaluation and feedback processes comparable to editorial focus groups, fine-tuning can recreate journalistic writing styles, editorial judgment and presentation formats that users find engaging and trustworthy. This optimisation extracts durable editorial value developed through publisher investment in journalistic talent and audience understanding, and therefore constitutes a distinct use of content that should require compensation. For example, in the case of AI Overviews, one of the objectives is almost undoubtedly keeping users within Google search surfaces in order to display ads to the user which means Google must provide editorially engaging and satisfying responses to the users. In keeping users on Google search surfaces to maximize the opportunities to render ads it will by definition reduce the click through rates on any attribution links that appear, and reduce traffic to publishers. An additional side effect is that fine-tuning improves subsequent grounding by enabling the model to better identify and prioritise the types of content users find most useful and engaging for specific tasks. We do not accept the use of our content for this purpose without our consent.

15. Further, if fine-tuning is excluded from the scope of the opt-in option, there is a material risk of circumvention. In future, we expect Google will rely more and more on fine-tuning rather than grounding. Google could well simply rely on frequent or continuous fine-tuning in a manner that effectively replicates grounding. Where publishers have decided not to opt in for use of their content for grounding for search generative AI features, the effectiveness of that would be significantly undermined if Google could fine-tune its models so frequently it no longer needed to ground its models to the same extent.
16. Enabling opt-in for fine-tuning will also facilitate content marketplaces, given the more granular preferences are able to be expressed, the more likely publishers are likely to be able to assess the value of opting-in for their content being used for a particular purpose. It will increase the options available, and more bespoke pricing, for buyers and sellers of content alike.
17. There is a suggestion from Google that is noted in the CMA's consultation document (paragraph 4.10(d)) that a decision not to opt in to fine-tuning would risk harming a publisher's ranking in general search. It is not clear to us what the CMA's view on this point is, why that would be the case, or what the nature and extent of harm to the general search ranking would be.

Recommendation 4 – Opting in should be on a per-product basis

18. To provide publishers flexibility and bargaining power in negotiations, publishers should be able to opt into different products separately – for example, AI Mode and AI Overviews. Different products have different implications for publishers: their referral rates, data provided for each and attribution format are different. Allowing publishers to opt-in on a per-product basis would also protect against Google's long-established practice of bundling products where it has market power with other products, such that publishers are forced to provide content for products they do not wish to.

19. This should not be difficult from a technical standpoint. For example, Microsoft provides AI specific reporting that enables publishers to assess how their content performs within AI-generated responses, demonstrating that per-product distinction is technically possible.¹¹

Recommendation 5 - Separation of Google's crawlers may be beneficial for the measure to be effective

20. As framed, Publisher CR(2) does not provide publishers with sufficient visibility over how their content is being accessed and used.¹² Crawler separation may avoid complex monitoring over the method through which publishers can express preferences, for example, publishers currently possess granular controls over how and when crawlers access content, including domain/subdomain/URL-level permissions, file-type and feature restrictions, crawl frequency, delayed crawls (on breaking news content) and location/geo blocking, and this all exists within existing publisher workflow and tools. Crawler separation is also necessary to facilitate development of content marketplaces. Finally, the current CR does not enable publishers to monitor or verify Google's compliance effectively. To make the remedy effective, the CMA should additionally require Google to separate its crawlers by purpose.

Access to data

21. The CR, as drafted, does not provide publishers with sufficient transparency as to how Google's crawler is accessing and using content. This is not mitigated by the transparency conduct requirement (Publisher CR(4)¹³). In particular, the combined effect of Publisher CR(2), (3) and (4) is that publishers will still not have any visibility over:

- a. the number of times Google's crawlers accessed content for different purposes (i.e. grounding, fine-tuning and/ or training), which is important for monitoring the value of content to Google for the purposes of any payment for content negotiations; and
- b. which dates and times content was accessed for different purposes. This could provide useful user insights as to how users are using and accessing content.

22. Therefore, even were Publisher CR(4) to cover more granular transparency requirements based on Google's usage of specific content – see our suggestions at paragraph 44 - 48 below – this would not alleviate the fact that publishers would still be reliant on Google to provide that information – which is obviously inappropriate in the context of any potential licensing negotiations.¹⁴ Crucially, it would also not be provided real-time – which it could be if Google's crawlers were separated and publishers could monitor crawler access as and when they accessed the site.

¹¹ Bing, *Webmaster Tools Help: AI performance in Bing Webmaster Tools*, available [here](#).

¹² CMA, *Consultation: Publisher Conduct Requirement* (28 January 2026), page 15, available [here](#).

¹³ CMA, *Consultation: Publisher Conduct Requirement* (28 January 2026), page 16, available [here](#); “4. Google shall: a. publish clear and detailed information explaining how Search Content is used for the training and grounding of its generative AI services and features; and b. ensure that the effect and scope of the controls described in paragraph 2 is transparent and can be well-understood by publishers.”

¹⁴ CMA, *Consultation: Publisher Conduct Requirement* (28 January 2026), page 16, available [here](#).

23. If the crawlers were separated, publishers would themselves have access to the information at 21(a)-(b), available on their own systems. Assuming they allowed Google to crawl content, they would be able to track how many times crawlers had accessed their site and for what purpose. The use of real-time data has been central to publishers’ editorial strategy: publishers use such data to identify which stories resonate with readers and to refine content accordingly.

Content marketplaces

24. DMG Media notes that the long-term development of sustainable value exchange is essential to ensuring that publishers receive fair and ongoing value for their content, without the need for continuous regulatory intervention. A functioning market mechanism that enables appropriate remuneration for high-quality journalism and editorial content would better align incentives across the digital ecosystem. Where publishers are properly rewarded for investing in original, resource-intensive reporting, there are clear benefits for society and democratic discourse. Conversely, if there is no effective means of compensating the production of high-quality content, commercial incentives will shift toward lower-cost alternatives, leading over time to a decline in quality and a race to the bottom.

25. As we set out in more detail below, many of DMG Media’s recommendations on the Publisher CR facilitate content marketplaces. We summarise this in the table below:

DMG recommendation	How this recommendation facilitates content marketplaces
Publishers should be able to opt into training for search generative AI features (see paragraph 9)	Greater granularity increases the feasibility of bespoke commercial models and expands the range of potential marketplace transactions. It enables product- and use-specific valuation and pricing.
Publishers should be able to opt in to fine-tuning, as a distinct activity (see paragraph 13)	
Opting in should be on a per product basis (see paragraph 18)	
Google should be required to separate its crawlers (see paragraph 20)	Crawler separation provides real-time visibility over access by purpose. This transparency underpins marketplaces and how publishers value content, enabling publishers to invest in high-performing content. Crawler separation also enables integration with third-party tools (e.g. pay-per-crawl models),
Publishers should have control over monetisation of their content through Search APIs (see paragraph 39)	API-level controls create the technical and commercial infrastructure for licensing access, enabling marketplace operators (Google or third parties) to channel demand into efficient marketplaces.
Transparency regarding data usage by Google (see paragraph 43)	Purpose- and product-level usage reporting enables accurate valuation and informed pricing. Transparency underpins sustainable licensing markets.
Product-level transparency over user engagement (see paragraph 49)	Product-level engagement data allows publishers to assess commercial impact on the consumer-end and negotiate appropriate remuneration.

26. A crawler remedy would facilitate the development of content marketplaces. Tools such as Cloudflare already enable publishers to manage and monetise crawler access directly. A publisher can choose to (a) block a crawler entirely; (b) block access unless payment is made; or (c) allow access.¹⁵ Under models such as Cloudflare’s proposed “pay-per-crawl” model, these options can be implemented within an integrated platform, with the controls over different AI crawlers linking directly to a paid-access option.
27. By contrast, it is far less clear how third-party tools could offer an equivalent monetisation solution where Google operates a single crawler and so cannot be blocked. In that scenario, access and downstream use are controlled within Google’s own systems, making it uncertain how a seamless independent marketplace model - where publishers can control and determine monetisation strategy across multiple AI companies in one portal - could function effectively. Google will always be a “special case,” making its interoperability with these models challenging and likely disadvantaging competing AI companies.
28. As above, if Google’s crawlers are separated, DMG will have access to real time usage data (paragraph 21 above). This transparency will underpin the effective development of content marketplaces as publishers will understand how their data has been used, in real time, and for what purposes, meaning they can assess the value of their content and price/license it with confidence.

Monitoring

29. The proposed CR is ineffective as it cannot be appropriately monitored or controlled by publishers and instead relies extensively on CMA monitoring. In particular, to meet its stated aim, it requires ongoing monitoring of:
- a. Google’s design of a sufficiently granular system to express preferences;
 - b. Google’s provision of an accessible, easy-to-use means to express those preferences, without friction;
 - c. whether Google has actually, in any given case, separated the combined data scraped by its combined crawler based on the expressed preferences; and
 - d. even if it has separated that data, whether it has recombined the data on the back-end.
30. Crucially, publishers cannot monitor Google’s compliance with 29(c) or (d). Publishers will not know if data scraped in any given instance was actually used in an AI Overview or in AI Mode, even if not attributed. They will only see that Google’s combined crawler has accessed their website, which could be for the purpose of Google Search or for Google’s generative AI features. They will have to trust Google to separate the data it has scraped based on their expressed preference. They are therefore reliant on the CMA, in perpetuity, to monitor Google’s compliance with 29(c) or (d). This does not give publishers much comfort: the indications are that Google often does not respect formal regulatory or legal processes.¹⁶ In addition, publishers will be reliant on a resource-constrained CMA to prioritise monitoring and enforcing of breaches of the conditions in 29(a)-(d) on an ongoing basis.

¹⁵ Cloudflare, *Introducing pay per crawl: Enabling content owners to charge AI crawlers for access (1 July 2025)*, available [here](#).

¹⁶ *Epic Games Inc v Google LLC*, No 3:20-cv-05671 (ND Cal, 28 March 2023); European Commission, *Commission sends preliminary findings to Alphabet under the Digital Markets Act (19 March 2024)*, available [here](#); European Commission Director-General for Competition and Directorate-General for Communications Networks, Content and Technology, *Commission opens investigation into potential Digital Markets Act breach by Google in demoting media publishers' content in search results (13 November 2025)*, available [here](#).

31. While there is still scope for circumvention, a remedy requiring the separation of crawlers reduces the risk by putting power back in publishers’ hands, such that they can monitor Google’s compliance themselves. How Google designs any mechanism for expressing preferences (i.e. 29(a) and (b) above) becomes less crucial, as publishers would be able to block Google’s crawler for AI services themselves through technical means.¹⁷ They would also not have to rely on Google retrospectively separating data scraped from a combined crawler (see 29(c) above) as there would be no need to separate data: each crawler would only scrape data for the relevant purpose. It is true that, even with a crawler separation remedy, the CMA would need to monitor that the crawlers are indeed separate – but we consider this is far easier to monitor: bots crawling for training, fine-tuning and grounding purposes typically behave very differently (for example in the frequency they access content).

32. We summarise these differences in the table below:

Factor which requires monitoring	Would the CMA have to monitor this with the current CR?	Would the CMA have to monitor this with a crawler separation remedy?
1. Google’s design of a sufficiently granular system to express preferences.	Yes – this would need to be subject to ongoing monitoring by the CMA.	Not as necessary – the market already provides bot management tools, which could be used to simply block Google’s crawlers. Publishers would no longer necessarily need to use robots.txt, as publishers could use third party tools to block the crawlers.
2. Google providing an accessible, easy-to-use means to express those preferences, without friction.	Yes – this would need to be subject to ongoing monitoring by the CMA.	Not as necessary, as preferences are no longer the be all and end all. Publishers could block crawlers directly, or using third party tools.
3. Whether Google has actually, in any given case, separated the combined data scraped by the bot based on the expressed preferences.	Yes – this would need to be subject to ongoing monitoring by the CMA. Publishers would also have no insight – so would not be able to alert the CMA to this issue.	Not as necessary, Google would not have a combined crawler, so there would be no need for the CMA to check it is separating the data on a case-by-case basis.

¹⁷ For example, through third party bot management tools.

<p>4. Whether it has recombined the data on the back-end.</p>	<p>This would need to be monitored, even where Google had separated the data.</p>	<p>Recombination would not arise in the same way: content that publishers have opted out of would not be scraped at all for the relevant purposes, removing the possibility of subsequent recombination.</p>
<p>5. Whether Google has separated its crawlers</p>	<p>This would not need to be monitored, as the crawlers would not be separate.</p>	<p>This would need to be monitored. However, given this is a structural change, we would expect monitoring to be light touch. In addition, crawlers for different purposes generally behave very differently (for example in the frequency they access content) so we consider this would be reasonably obvious.</p>

33. Competition authorities have long recognised the advantages of structural remedies, particularly in the context of merger control, where they are generally preferred for their durability and ease of enforcement. The CMA should not depart from this well-established approach in the present context.

Costs of crawler separation

34. Google claims it would cost over £150 million per year to separate its crawlers, compared to £3.5 million per year for the proposed conduct requirements.¹⁸ We do not agree with Google’s approximation of the costs. First, Google already operates at least twenty-four separate crawlers.¹⁹ This suggests it is not beyond the realms of possibility for it to operate twenty-five. In addition, most other AI companies have separate crawlers for grounding, fine-tuning and training their AI models, suggesting this is the optimal, cheapest and most efficient model in the parallel situation where AI companies are not trying to leverage market power from Search.²⁰

Impact on server load and end users

35. We understand it may have been argued that requiring Google to separate its crawlers would increase server load and may have adverse impacts on user experience. We do not consider that Google splitting its crawlers would have a material increase in server load for publishers or adverse effects for end users. The consultation document does not provide evidence demonstrating that the operation of a distinct crawler for AI purposes would generate incremental traffic beyond the ordinary levels associated with automated web indexing. In practice, a separate crawler would constitute one of many automated agents routinely accessing publicly available webpages.

¹⁸ CMA, *Consultation: Publisher Conduct Requirement* (28 January 2026), para 5.26, available [here](#).

¹⁹ Google Developers, *Overview of Google crawlers and fetchers (user agents)*, available [here](#); Including 11 [common crawlers](#), 5 [special case crawlers](#), and 8 [user-triggered fetchers](#) <accessed 12 February 2026>

²⁰ Playwire, *The Complete List of AI Crawlers and How to Block Each One* (8 December 2025), available [here](#).

6. Moreover, because Googlebot currently combines crawling for Search and AI purposes, it is typically permitted by publishers in circumstances where a crawler dedicated solely to AI uses might not be. Although Googlebot generates substantial traffic, it is comparatively rarely disallowed via robots.txt. If AI-related crawling were separated, publishers would be able to exercise more granular control and may choose to restrict AI crawling while continuing to permit Search indexing. In that scenario, total crawling traffic may in fact decrease relative to the status quo, rather than increase.

Recommendation 6 – Google could use publisher content via overseas datasets

36. The current CR states that Google “*shall not attempt to circumvent any publisher’s choice to withhold its Search Content by acquiring that content through other sources*”²¹. The interpretive notes qualify this however, stating it would be permissible for Google to acquire such content “*through open-source datasets, where these datasets have obtained content legally, given the nature of the sources*”.
37. We are concerned that many prominent open-source datasets, such as Common Crawl, have been assembled through large-scale web scraping without express consent. Common Crawl was founded in 2007 by a former Google engineer as a non-profit organisation and has been supported in part by funding from AI companies.²² Datasets of this kind can function as an intermediary that allows AI developers to access web content at scale without directly obtaining it from publishers. Once content is incorporated into open-source datasets, it is extremely difficult to remove in practice.²³ Many widely used datasets are derivatives of other open source datasets meaning publisher content can persist across multiple downstream datasets, for example, C4 (Colossal Clean Crawled Corpus), was developed by Google using a filtered version of Common Crawl, and many other derivate open source data sets exist such as Refined Web, LAION and FineWeb. Content may also enter these datasets indirectly where material is republished outside publisher-controlled domains, such as through distribution partners or third-party republication, and subsequently scraped without consent. In addition, some open-source datasets are themselves assembled from other datasets rather than direct scraping. Open source datasets are widely replicated, redistributed globally, and incorporated into downstream training corpora, such that removal from an original dataset does not eliminate downstream copies. As a result, it is exceptionally difficult to determine whether content contained within any public open-source dataset was lawfully obtained. If such datasets are treated as “lawfully” obtained in jurisdictions with more permissive copyright regimes, and are then used to train or ground AI systems deployed in the UK, this could create a material loophole in the proposed framework.
38. We consider this risk could be addressed by either:

²¹ CMA, *Consultation: Publisher Conduct Requirement* (28 January 2026), page 16, available [here](#).

²² https://www.aol.com/articles/common-crawl-accused-feeding-paywalled-184021524.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAJ-g37HZO4XbArSUhw_qJBtcYhaCPoP9nkq6cqmQIEhIF6BPXcSfYdGiz5--NDKsCO-yMNEVSN-OVG6IG_-Oss6TRAU7pBzn5KoNmUXoWUKYKqogI_aPRUTnT_A3pqmrbInfUvSyiXnJVKjdPHOWYQUHhDDAZ_1HjgcVcNQ61hG

²³ <https://www.theatlantic.com/technology/2025/11/common-crawl-ai-training-data/684567/>

- a. clarifying that “legally obtained” means lawful under UK copyright standards - i.e. that the dataset would have been lawful had UK law applied at the point of collection - thereby preventing reliance on more permissive overseas regimes to circumvent the conduct requirement; and/ or
- b. at a minimum, requiring Google to exercise reasonable and proportionate due diligence to satisfy itself that the datasets relied upon for UK-facing services do not contain content obtained in a manner inconsistent with UK copyright law.

Recommendation 7 – Google monetisation of its Search APIs for third parties to use in grounding

39. The CR regulates Google’s direct use of publisher content for its own AI services, including grounding in products such as Gemini. However, it does not address Google’s ability to monetise or license access to that same content to third parties via its Search APIs, including where those third parties use the content for AI grounding purposes. This creates a material gap in scope.
40. In particular, the CRs do not prohibit Google from either directly or indirectly commercially supplying access to its search index or related APIs for AI use cases, even where those uses rely on publisher content. Google Search functions as a grounding layer derived from publisher content and user interaction. Third-party AI developers can programmatically access, via the Google Search API, ranked URLs, snippets and summaries, freshness and popularity signals. Google is uniquely positioned to provide these signals at scale given its extensive visibility across the web, including via Google Analytics across publisher websites. While Google may claim to limit high-volume API access (perhaps to reduce direct competition with its own AI features), intermediaries such as [CONFIDENTIAL] enable third parties to access Google search results for AI grounding at scale, with Google compensated for providing API access while publisher content underpins the underlying outputs. CRs should therefore make clear that Google may not license, sell or otherwise provide access to publisher content via Search or grounding APIs for third-party AI purposes, unless subject to equivalent controls and protections.
41. If effective controls are introduced over the use and sale of publisher content through Google’s Search API, this would help enable the development of content marketplaces. For example, Microsoft allows publishers to control whether their content is used through its grounding tool and API. It has built on this infrastructure, which is a hub of buyer demand for content, to develop a content marketplace that connects publishers with parties seeking licensed content.
42. If Google were required to offer similar controls over how its API uses publisher content, this could create the foundations for comparable marketplace models - whether operated by Google itself or by third parties - allowing demand for licensed content to be leveraged in an efficient way.

Transparency and attribution – CR(4), CR(5) and CR(6)

43. We agree that a transparency conduct requirement is necessary and support the proposed obligations regarding (a) data on how Google uses publishers’ content (Publisher CR(4)²⁴) and (b) data on user engagement with content (Publisher CR(5)²⁵). However, we consider that additional measures are required for the transparency CR to be fully effective.

Google’s use of content

44. In relation to Google’s use of content, the conduct requirement refers to the publication of “clear and detailed information” explaining how Search Content is used. However, this obligation appears to operate at an aggregate level across all publishers, rather than requiring publisher-specific reporting. In particular, it does not oblige Google to inform individual publishers how often their content has been used for the fine-tuning, training or grounding of generative AI features.

45. Such granularity is essential if publishers are to make informed choices and properly assess the value of their content. As noted above, we consider that a structural remedy - such as crawler separation - would address this concern to some extent by providing publishers visibility over Google crawler visits on their own sites. However, if the CMA does not intend to adopt a crawler separation remedy – it should ensure publishers have access to the same data as if they had.

46. In respect of how Google uses content, we consider Google should be required to provide the following:

- a. **Purpose level usage reporting, disaggregated by publisher:** including clear separation between training, fine-tuning and grounding. This is necessary to enable publishers to understand precisely how their content is being exploited and to assess the relative value and implications of each type of use. As above, with a crawler separation remedy, publishers would have access to this data in any case.
- b. **Product-level usage reporting:** showing which specific AI products or features used our content. This should cover both search generative AI and the broader generative AI services by specific product (AI Overviews, AI Mode, Gemini apps, Vertex API etc.) This is necessary to enable publishers to understand precisely how their content is being exploited and to assess the relative value and implications of each type of use.
- c. **Disclosure of zero-impression usage,** i.e. where content is used for grounding without attribution. This is critical because such uses generate value for Google without providing traffic or brand visibility

²⁴ CMA, *Consultation: Publisher Conduct Requirement* (28 January 2026), page 16, available [here](#).

²⁵ CMA, *Consultation: Publisher Conduct Requirement* (28 January 2026), page 16, available [here](#); “5. Google shall provide publishers with clear and detailed metrics on user engagement with their Search Content where it is used in its search generative AI features”.

in return. This would also enable publishers to assess Google's compliance with the attribution CR (CR(6)).²⁶

- d. **URL-level access and sourcing transparency:** Identification of the specific publisher URLs accessed, including the timing and frequency of access, the relevant purpose and product for which the content was used, and the crawler or alternative data source through which the content was obtained (where access did not occur via a publisher-facing crawler). This level of transparency is necessary to enable publishers to assess the scale and commercial value of the use made of their content and also identify illegal sources of information.

47. The disclosure of data on Google's use of content and how users engage with content is essential to the effective functioning of content marketplaces. By understanding how their data has been used, in real time, and for what purposes, publishers can assess the value of their content and price and license it with confidence. Transparency is therefore a precondition for a sustainable and efficient marketplace. As above, this would be better in real time.

48. Greater visibility over usage would also support investment and innovation. When publishers can see which types of content generate value, they are better able to allocate resources and produce material that meets demand. This dynamic is well established in digital publishing: for example, media organisations such as DMG Media use real-time performance data to identify which stories resonate with readers and to refine their editorial strategy accordingly. Comparable transparency in content usage would similarly strengthen incentives to create high-quality material.

User engagement data

49. In relation to user engagement data, this should be provided on a product-by-product basis: Google should provide clear, product-level referrer data identifying which specific AI product or feature generated a referral, as current referrer information does not offer this granularity. Without this visibility, publishers such as DMG Media cannot reliably measure traffic impacts, evaluate the performance of different interfaces, or make informed strategic and commercial decisions about participation and licensing. Similar to the usage data, this transparency also provides the underpinning for a sustainable marketplace.

Attribution

50. In relation to the attribution CR (Publisher CR(6)²⁷), this is impossible for publishers to rely on without stronger transparency of how Google has used data, as publishers will not know when their content has been used but not been attributed. The CR is also drafted at a high level and uses the vague term "sufficiently" attributed.

²⁶ CMA, *Consultation: Publisher Conduct Requirement* (28 January 2026), page 16, available [here](#). "6. Google shall: a. Take reasonable steps to ensure that Search Content is sufficiently attributed when used in search generative AI features; and b. Publish clear and detailed information explaining the steps it takes to: (i) Ensure that Search Content is sufficiently attributed in search generative AI features; and (ii) Ensure and measure the factuality of search generative AI features"

²⁷ CMA, *Consultation: Publisher Conduct Requirement* (28 January 2026), page 16, available [here](#).

We consider, at a minimum, it should be amended to state any attribution should be in close proximity to the content and have appropriate prominence and visibility.

Monitoring

Independent, third-party oversight

51. We consider Google’s six-monthly compliance report should be produced or, at a minimum, audited by an independent third-party in order to improve trust.²⁸ This is particularly the case given the CR is designed to “address [...] a lack of trust by publishers”.²⁹ This would bring the compliance report in line with the initial baseline compliance audit, for which the CMA consider there is “intrinsic value” in being produced by an independent third party.³⁰
52. Independent oversight would ensure that any testing is designed to assess real-world outcomes, including through adversarial testing methodologies where appropriate. This would anchor the assessment in objective principles, designed by specialist third party auditors, rather than Google defining the parameters and metrics by which its own compliance is judged. It would help to identify and correct any inadvertent blind spots in Google’s testing methodologies, including those that may have arisen from internal groupthink. It would therefore enable evidence-based refinement of the CRs over time, to the benefit of both Google and publishers where aspects of the regime prove ineffective or require adjustment.
53. This approach would be consistent with established market practice in adjacent markets. For example, record labels rely on independent third parties, such as Vemillio, to monitor and verify usage of their content on YouTube. While record labels may possess greater bargaining power given their relative scale, the underlying principle is the same: independent scrutiny is necessary to provide confidence in reported usage data. Publishers lack both the visibility and leverage to scrutinise how their content is being used within Google’s AI systems and therefore require comparable third-party oversight.
54. Further, we consider that an independently produced compliance report would prevent Google from deliberately evading regulatory oversight, as it has done on previous occasions. A stark instance of this was in the recent US litigation: in *Epic v Google*, Judge Donato ruled that internal chat evidence was “lost with intent” to “hide the ball”.³¹ Another instance is the DMA, where Google is now under three separate investigations for breaching its obligations.³²

²⁸ CMA, “*Consultation: Publisher Conduct Requirement*”, p. 42, available [here](#).

²⁹ CMA, “*Consultation: Publisher Conduct Requirement*”, p. 42, available [here](#).

³⁰ Ibid.

³¹ CNN, “*Federal judge rules Google tried to ‘hide the ball’ by deleting chat logs in a big antitrust case*”, 29 March 2023, available [here](#).

³² European Commission Director-General for Competition and Directorate-General for Communications Networks, Content and Technology, *Commission opens investigation into potential Digital Markets Act breach by Google in demoting media publishers’ content in search results* (13 November 2025), available [here](#); European Commission, *Commission opens proceedings to assist Google in complying with interoperability and online search data sharing obligations under the Digital Markets Act* (27 January 2026), available [here](#). European Commission, *Commission sends preliminary findings to Alphabet under the Digital Markets Act* (March 19 2025), available [here](#).

55. We also submit that, while Google will have to pay a third-party to carry out this compliance report, this cost will be insignificant relative to Google's scale and financial resources.

Format of report

56. We consider the six-monthly report should be structured and, in addition to the information addressed at 4.74 of the Publisher CR³³, should also include: (a) the details of new features it is releasing to comply with the CR, (b) how it is making third parties aware of any new features or changes to features related to publisher controls and (c) how many publishers have opted-in for the controls. The latter be an indicator of effectiveness/ whether any changes need to be made to the CR.

Payment for content conduct requirement (CR)

57. The CMA has stated it does not intend to take further action on payment for content for at least 12 months.³⁴ Respectfully, this is not fast enough. The CMA first considered payment for content proposals, at length, in its Digital Advertising Market Study in July 2020 and its Joint Advice with Ofcom in November 2021.³⁵ We are now in 2026 and the CMA has said it will not take action until 2027. Further delay would materially prejudice publishers given current market dynamics. The CMA's crawler remedy is helpful but, even if it achieves its objectives in full, it clearly cannot address the unfair value exchange that exists in both general search and search generative AI products, nor can it address the historic and continuing unlicensed use of news content to train AI models. The consultation does not set out a clear rationale for deferring the payment for content measures.

58. Given the pace of AI development, it may soon become commercially unrealistic for publishers to decline to opt in to AI Overviews and AI Mode, however unattractive the terms. These products may become the functional equivalent of Google Search itself - critical to publishers' visibility. If that occurs, publishers will once again lack any meaningful bargaining power and will be subject to Google unilaterally determining the value, if any, attributed to content used in these AI interfaces. Therefore, while we strongly support controls over use of publishers' content, we consider this should be developed alongside a robust CR for ensuring fair payment to publishers, as was explicitly provided for in the Digital Markets, Competition and Consumers Act 2024.³⁶

59. Even if the CMA considers that now is not the appropriate moment to consult formally on the details of a CR relating to payment for content, it should at a minimum begin developing the detailed design of such a requirement, including through engagement with stakeholders, so that it is ready to act if necessary. The worst-case outcome would be for the CMA to conclude in 12 months' time that the current package of CRs is

³³ CMA, "Consultation: Publisher Conduct Requirement", para 4.74, available [here](#).

³⁴ CMA, Will Hayter, *Improving the way Google delivers search services in the UK* (28 January 2026) available [here](#).

³⁵ CMA, "Online platforms and digital advertising: Market Study final report", available [here](#). Ofcom, CMA "Platforms and content providers, including news publishers: Advice to DCMS on the application of a code of conduct", available [here](#).

³⁶ Digital Markets, Competition and Consumers Act 2024, s20(2)(a) and s38(1).

ineffective, only then to begin designing a remedy - resulting in a further year (or more, given the likely implementation period) of delay that publishers can ill afford.

Fair ranking

Explicit prohibition of SRAP

60. We welcome the CMA's fair ranking CR.³⁷ As the CMA will know from previous submissions, DMG Media has itself been a victim of Google's arbitrarily applied ranking policies.³⁸ Google's 2024 Site Reputation Abuse Spam Update, now the subject of an EC investigation under the Digital Markets Act, forced Daily Mail to remove articles with retailer discount codes.³⁹ This was an arbitrary policy which immediately reduced a valuable revenue source to zero. The result has been that users are no longer given guidance from search results for answers to questions such as "what are the best washing machines." Instead, they are more likely to ask AI Mode or AI Overviews, which may well have been Google's intention.
61. The CMA itself appears to recognise that this conduct was an anticompetitive application of Google's market power, as evidenced by its repeated references to the practice throughout the CR document.⁴⁰
62. However, we are concerned that the proposed CR does not require Google to reverse its "site reputation abuse" policy. Although the policy is referenced in the consultation document, there is no explicit prohibition of the practice in the CR itself or in the accompanying explanatory notes. Instead, the proposed approach relies on a complaints mechanism, whereby publishers may complain to Google, which then, in its own discretion, provides summaries of complaints to the CMA, which then can request a report from Google assessing whether a policy is having a material adverse impact and setting out the steps Google proposes to take to address it. Not only does this appear to be a lengthy process, even at the end of that process it does not follow from the CR that the CMA can bind Google to any particular outcome if it disagrees with Google's assessment or the steps Google has proposed. The CMA would likely have to rely on its residual powers to investigate under the DMCCA, such as a pro-competitive intervention investigation, which again would add months or even years to the timeline of resolving any complaints.
63. This provides little reassurance. DMG Media has already raised concerns about the site reputation abuse policy in multiple prior submissions, as have many publishers. Given the CMA's apparent implicit recognition of the competition concerns associated with this practice, the CR should include an explicit prohibition or clear requirement to withdraw the policy, rather than relying solely on a lengthy and uncertain *ex post* complaints process.

³⁷ CMA, *Consultation: Fair Ranking Conduct Requirement* (28 January 2026), available [here](#).

³⁸ DMG Media, *Proposed decision on the CMA's strategic market status investigation into Google's general search services. Response from DMG Media*, para 51-52, available [here](#).

³⁹ European Commission Director-General for Competition and Directorate-General for Communications Networks, Content and Technology, *Commission opens investigation into potential Digital Markets Act breach by Google in demoting media publishers' content in search results* (13 November 2025), available [here](#).

⁴⁰ CMA, *Consultation: Fair Ranking Conduct Requirement* (28 January 2026), available [here](#).

64. Even if the CMA were to allow the site reputation abuse policy to continue, it is clear that it could be exercised in a significantly more proportionate way that does not prohibit news publishers from using marketing tools, which is something newspapers have always done since long before Google existed. For example, there could be a clear threshold for the policy that allows these marketing tools to be used as long as they represent a minority of the website's content. Reaching a reasonable solution quickly is even more pressing given that Google's practices related to AI Overviews and AI Mode are driving down advertising revenue streams for publishers.

Transparency

65. We welcome the aim of more transparency in how Google ranks content. However, the notice obligation in Fair Ranking CR(5)⁴¹ does not go far enough. To be effective, the CR should specify minimum notice periods. This should be a minimum one month for straightforward changes, and six months for major changes where businesses may need to make technical adaptations to comply with or benefit from the changes. If the CMA does not consider it appropriate to include this expressly in the text of the CR, it should at a minimum make clear in the interpretative notes that this is the approach it would ordinarily expect Google to adopt, absent compelling evidence that it would be inappropriate in a specific situation.

Complaints mechanism

66. Publishers such as DMG Media consistently have difficulty engaging with Google within their existing complaints process, giving rise to significant costs.⁴² Therefore, it is necessary to impose clear, binding conditions to ensure complaints are handled in an accessible and timely manner.

67. We do not consider the current CR in Fair Ranking CR(6) is specific enough.⁴³ It simply refers to a "clear and accessible" process. This does not provide any guarantees as to the level of performance Google will provide in respect of its complaints process. In particular, the CMA should ensure:

- a. Google commits to a target timeline for responding to complaints (which is no longer than 14 days) and report their performance against that timeframe;

⁴¹ CMA, *Consultation: Fair Ranking Conduct Requirement* (28 January 2026), page 19, available [here](#); "5. Google shall provide publicly: a. sufficient information about how it undertakes the matters described in paragraph 2; and b. sufficient notice and information in advance of any material changes to the key ranking criteria, policies and procedures that affect the matters described in paragraph 2."

⁴² AITO's response to the invitation to comment; Checkatrade's response to the invitation to comment; DMG Media's response to the invitation to comment; European Publishers Council's response to the invitation to comment; Impress, the Independent Media Association and the Media Reform Coalition's response to the invitation to comment; Lovehoney's response to the invitation to comment; Network N Media's response to the invitation to comment; News Media Association's response to the invitation to comment; Platform Leaders' response to the invitation to comment; Professional Publishers Association's response to the invitation to comment; Public Interest News Foundation's response to the invitation to comment; BBC, Premier League and Sky's response to the invitation to comment.

⁴³ CMA, *Consultation: Fair Ranking Conduct Requirement* (28 January 2026), page 19, available [here](#); "6. Google shall provide a clear and accessible process for handling complaints from publishers in relation to its policies and procedures related to any of the matters described in paragraph 2, which allows publishers to indicate any potential adverse impacts on the functioning of any market in the United Kingdom in which they are active"

- b. Google provides the outcome of any complaint with detailed reasons and evidence and gives the complainant an opportunity to respond on the points raised;
- c. all complainants who are unsatisfied with the outcome of their complaint have the ability to appeal to an independent adjudicator. This is currently limited to where Google has manually excluded the publisher from Google's general search index (Fair Ranking CR(9)).⁴⁴ However, this should be the case for all complainants or – at the very least – when any part of a publisher's website (e.g. a directory or a subdomain) has been excluded.

Concluding remarks

68. DMG Media welcomes the CMA's leadership in addressing Google's entrenched market power in general search and the emerging risks associated with generative AI integration. The proposed CRs represent an important step toward restoring fairness, transparency and accountability in digital markets. However, as currently drafted, the package risks being insufficient to address the scale and urgency of the competitive harm facing UK publishers.

69. In particular, a number of the proposed measures rely heavily on ongoing monitoring, broad standards, or ex post complaints processes, rather than clear, enforceable and structurally robust obligations. Given Google's scale, technical sophistication and track record of resisting regulatory constraints, remedies must be durable, objectively verifiable and capable of operating effectively without continuous regulatory intervention.

70. The CMA has a historic opportunity to set a global benchmark. To ensure the regime is effective in practice, the CRs must provide genuine publisher control, meaningful transparency, clear accountability and a credible pathway to fair remuneration. Without strengthening the current proposals, there is a real risk that competitive harm will continue while enforcement processes play out over extended timelines that publishers cannot afford.

71. Finally, we are concerned that the CMA is adopting unduly burdensome interpretations of the words "effective" and "proportionate". These words do not mean that the CMA must always adopt the weakest remedy or a remedy that Google says it can live with. Where, as in this situation, a weaker remedy carries higher risks of ineffectiveness, the CMA is fully entitled to adopt the stronger version.

25 February 2026

⁴⁴ CMA, *Consultation: Fair Ranking Conduct Requirement* (28 January 2026), page 19, available [here](#); "9. In addition to the complaints process described in paragraph 6, Google shall also publish an alternative dispute settlement mechanism for publishers that have been manually excluded from Google's general search index."