

## GMG response: CMA's proposed Phase 1 CRs for Google search

Executive Summary	1
Key Areas of Concern	1
<b>PUBLISHER CONTROLS CR</b>	<b>2</b>
Granularity of controls (6.2a)	2
*Fine-tuning:	3
Page level controls (6.2b):	4
Performance and engagement information on a 'per-feature' basis (6.3a)	5
How search content is attributed in search generative AI features (6.4b)	7
Third-party crawling:	7
Accelerated timeframe:	9
<b>FAIR RANKING CR</b>	<b>11</b>
Is the FR scope correct (6.2)	11
Transparency	12
Non-distortion and complaints	12
<b>USER CHOICE CR</b>	<b>12</b>
Frictionless opt out:	12
Gemini inconsistency:	12
<b>AUDITING MEASURES (for all CRs)</b>	<b>13</b>
<b>SUMMARY OF PROPOSALS</b>	<b>13</b>

### About GMG

Guardian Media Group (GMG) is one of the UK's leading commercial media organisations and a British-owned, independent, news media business. GMG is the owner of Guardian News & Media (GNM), which is the publisher of theguardian.com and the Guardian newspapers, which have received global acclaim for investigations including persistent investigations into phone hacking amongst the UK press, the Paradise Papers and the Pegasus Project. As well as being the UK's largest quality news brand, the Guardian has pioneered a highly distinctive, open approach to publishing on the web and it has achieved significant global audience growth over the past 20 years.

### Executive Summary

The proposed Conduct Requirements (CRs) for Google represent a critical juncture for the UK's digital economy. We know that regulators around the world are looking at the CMA - and there is a high chance that these 'remedies' become a global 'solution' for the next few, critical, years. The stakes are high.

The underlying intent of these requirements is sound and comprehensive. However we are concerned that the mechanisms for their enactment will all take place within the Google

ecosystem raising issues around visibility and accountability. This creates the risk of material gaps that could undermine the objective of giving publishers meaningful control over their content.

## Key Areas of Concern

- **Fair Ranking between features:** In the time that has elapsed since the start of the CMA process, we have seen Google relegate search within its home page and promote AI Overviews. There is a pressing need for fair ranking between features as well as within them to ensure that the choice these new controls provide is meaningful.
- **Publisher Control:** We maintain that a split crawler would give more real time visibility and clarity of crawler purpose. We also challenge the arguments around costs. To strengthen the recommended solution, we recommend the CMA mandate separate, defined opt-in controls for all Google services including search, grounding, fine-tuning, training and all other features specified by name.
- **Per feature data:** Control can only be meaningful if publishers have granular data to make informed decisions. Our preference is that data should be split between features and include citation and attribution information as well as clicks, CTR and impressions.
- **Transparency and Auditing:** Robust, ongoing 3rd party auditing is necessary to ensure that opting out does not impact rankings and that data is not recombined in the background.
- **Timetable for Implementation:** Google is moving rapidly to cement its search advantages into AI integration. The CMA's timeline therefore needs to accelerate, and is essential for publisher controls in particular.

## PUBLISHER CONTROLS CR

### Granularity of controls (6.2a)

The CMA's proposals leave it to Google to decide the technical means by which separate controls are achieved. Our experience to date of Google's AI controls is Google Extended (GE). The CMA notes that the current definitions in GE are ambiguous and that Google has been unwilling to clearly define 'broader Gen Ai services' in a granular manner

We welcome the CMA's recommendation (1a(ii)) that Google "*publish clear and detailed information describing the scope of the GE controls. This should include a description of GE and its overall purpose, the key exceptions and limitations to its scope and an explanation of the publisher content GE cover.*"

However, 'overall purpose' is too vague. Meaningful control can only be achieved if **publishers are able to opt in on a per service basis - with each service individually specified and its purpose narrowly defined - for both existing and future products i.e.**

- **Training**
- **Fine Tuning \***
- **Grounding**
- **Search**
- **Other named AI features**

This is important for fair bargaining as each use will have a different substitutional effect and therefore a different value. This granularity should extend to Google Extended as well as the Googlebot.

The CRs require Google to submit its implementation plan within a month and then enact it within 5 months. Although there is a requirement to ‘engage constructively with the CMA and third parties’ during that period, it is unclear whether there is a redress mechanism if publishers feel the controls have been ‘cherry picked’ (4.18a) or are unclearly defined - a risk that the CMA itself notes.

Further recommendations include:

(i) specifying that **controls should be standardised** so that publishers do not have to manage different technical signals such as bespoke tagging instructions for every feature Google launches.

(ii) that Google transparently disclose **new AI products (and their controls) 30 days before any launch (whether trial or public)** to allow publishers to set controls. We recommend this sits alongside the opt in options on the Search console (see iii).

(iii) that the **opting in process is clear and simple**. We would suggest it is a feature of the search console and with clear toggles that operate in a similar way to how cookie consent is given. This will benefit smaller publishers with limited technical resources.

\*Fine-tuning:

**Substitutional product threat:** The CMA appears to accept Google’s claim (4.10c) that using content to fine tune models has no harmful impact on publishers despite that same fine tuning keeping outputs up to date. This clause demonstrates the clear value of publisher content at every stage: training, fine-tuning, and grounding - but in particular in keeping outputs accurate. Given that news publishers are trusted sources across a wide range of queries, this would suggest that their content **is** being used to create substitutional products - invalidating the argument that it causes ‘no harm’ to publishers.

**Eroding value of RAG content:** We are seeing a trend<sup>1</sup> towards more frequent, bespoke fine-tuning to improve model accuracy, the impact of which is to improve the recency of underlying LLMs, i.e. their ability to include current information in their responses. This in turn risks devaluation of emerging and valuable RAG markets for news content.

**Link between training and ranking:** Google also says that “*Allowing publishers to opt out of the fine-tuning of models underlying search generative AI features would be*

---

<sup>1</sup><https://techcommunity.microsoft.com/blog/azure-ai-foundry-blog/fine-tuning-together-how-our-developer-community-is-shaping-the-future-of-custom/4432495>

*counterproductive...*” (4.10d) This suggests an ongoing link between use in AI overviews and ranking in search which could negate meaningful publisher choice.

**Google’s incentive to merge:** The section goes on to say that “*If Google were to try to maintain the hard distinction between generative and ranking models, it would likely need to develop and maintain duplicative models to achieve the same quality improvement.*”

“*Maintain[ing] the hard distinction,*” suggests that this is the current situation and therefore keeping AI training and search ranking separate is possible. Merging them is a choice not an inevitability but the sentence implies that the direction of travel is closer integration. As is noted in the aims of the CRs, publishers should be able to opt out of generative fine tuning without it affecting their search ranking, otherwise Google’s overwhelming dominance in search risks transferring into the AI market.

Given this ambiguity, **GMG strongly recommends that fine-tuning is separated out as a control in its own right**

Page level controls (6.2b):

We welcome the CMA’s recommendation that within general search, an effective control should allow publishers to opt out at both directory-level and page-level (4.22).

**Argument for page level control in GE:** The logic expressed in 4.33 for page level controls in search is equally valid for training, fine tuning and grounding as different data sets have different market values and permissions. We question the premise that this would be ‘onerous’ to extend to Google Extended given that its competitors, including Microsoft, operate page level controls.

**Technical feasibility and precedence:** Google already processes a number of parameters read from the page meta-data: noindex, nofollow, no archive, no snippet, etc. There is also “x-robots-tag” - a HTTP header that works with non html content (like images, videos, pdf files etc). Extending a similar tag-based mechanism to GE and the new controls request is **technically feasible and proportionate compared to the harm of total site opt-outs**. The argument that is taxing to do via robots.txt ignores the other options.

Performance and engagement information on a ‘per-feature’ basis (6.3a)

Google has publicly stated its commitment to “supporting a healthy news ecosystem.”<sup>2</sup> Providing more granular data would be a concrete step towards supporting that aim from an organisation that prides itself on being data driven<sup>3</sup>.

The reasons cited by Google in the CRs (4.43 a,b) for not giving performance data on a per feature basis are that they would ‘*not be futureproofed*’ and that this has ‘*not been historically provided at a per feature level in non-gen Ai search results*’. The first is in Google’s gift to resolve; the second is perpetuating an existing problem.

---

<sup>2</sup> <https://committees.parliament.uk/writtenevidence/128498/html/>

<sup>3</sup> <https://cloud.google.com/resources/building-a-data-driven-culture>

## What are the actionable benefits?

**Levelling up the information asymmetry:** The data is crucial for two key reasons:

**1/ Improve the bargaining position of publishers:** We are disappointed that CRs enabling ‘fair negotiating’ have been moved into post 2027 CRs. A mitigation would be mandating granular data that may allow for fairer negotiations in the interim.

**2/ Assess impact and optimise content accordingly:** We note the repeated use in the CRs of the phrase ‘Quality Clicks’ (6.3b). This refers to a metric introduced in a [Google blog](#) from August 2025 which frames the loss of direct publisher traffic from AI overviews as being offset by this new, Google-determined metric. GMG challenges the premise of this new metric and questions the efficacy of framing 4.46-4.49 through this prism. Publishers value all human traffic, regardless of whether audiences arrive at our site through endeavour or more casual browsing. Determining quality of clicks is for content creators, not an intermediary.

We therefore welcome the CMA’s suggestion that publishers be given the data to “calculate this metric themselves” and independently verify “quality”.

**What ‘good’ looks like - comparison with Ophan:** The CR document asks for further evidence of the benefits of Google providing performance and engagement information. It may be helpful to compare the paucity of data for how content performs in AI summaries with ‘Ophan’, the Guardian’s inhouse tool developed to help editors, journalists and commercial teams understand how content is performing.

Attached is a single page from Ophan showing the depth of information we gather on every single article published on our website. All staff have access to Ophan - it is viewed as a critical tool in providing the feedback needed to improve our journalism and story choices for the benefit of our audiences. The sizable resource spent developing this tool reflects the value of granular data in driving effective business decisions as a news organisation.

We would also point the CMA to Microsoft’s recent decision to pro-actively publish some AI data for its Bing crawler.<sup>4</sup> It is a start but unfortunately not a blueprint. The inclusion of citation data is useful, but the omission of click data, and by extension CTRs makes it impossible for publishers to understand the meaningful value being cited delivers. Without data from the platforms we have to rely on third party tools or extrapolate from other signals the impact of AI generated summaries on CTR. Our observation is that when users are served an AI overview they are far less likely to click through to the source.

In addition to clicks, impressions and CTR, we suggest the following data points be mandated by the CRs.

**(i) Splitting out services in Search Console:** We recommend that the CMA mandate feature-level reporting parity with existing Discover surfaces in the Search Console.

---

<sup>4</sup> <https://blogs.bing.com/webmaster/February-2026/Introducing-AI-Performance-in-Bing-Webmaster-Tools-Public-Preview>

Publishers require **fully disaggregated data for every AI feature** (AIO, AIMode as well as any future features).

**(ii) CTR comparison between search and AIO:** Data that allows publishers to clearly see the **difference in CTR for blue links that appear in traditional search v blue links that appear beneath an AIO**. The ability to benchmark AIO data against search data is a prerequisite to publishers making an informed decision about whether or not to opt out of AIO.

**(iii) Url referrers:** Google should **reintroduce url referrers** as a metric to publishers' meaningful visibility of user journeys on a per article basis. It should be noted that paid traffic (Google Ads) still provides detailed keyword data to advertisers.

In order to 'future proof' the data requirements as Google moves further towards AI features in search we would also recommend including **attribution data points**.

**(iv) Total Citations** - as well as showing the total citation (number of citations that are displayed as sources in AI generated answers), these should be broken down into categories of answers or key words used by users. Understanding the grounding queries that lead to publisher content being cited would put AI answers on a par with traditional search.\*

**(v) Accredited citations** - this shows the number of unique pages which are accredited as a source in AI generated answers.

**(vi) Accredited citations as a % of grounding score** - this would help to understand the percentage of the original text in the AI answer that triggers an attribution.

\* It would be helpful to stipulate that the process of caching should not be used to skew the data. This is where an LLM periodically caches its search index (perhaps once a day) and then uses that cached data to ground multiple answers in AIO or AIM. In this manner a single crawl can generate multiple answers.

## How search content is attributed in search generative AI features (6.4b)

We are pleased to see the CMA recognise the difficulty publishers face in identifying how their content is attributed and any factual inaccuracies, given the non-canonical nature of AI overviews. Monitoring at scale is impossible without Google's assistance. However we are concerned at the lack of specificity in the wording of the CRs. Neither publishers nor the public are well served by the current limited or missing attribution - and 'sufficient attribution' or 'reasonable steps' give no guarantee of improvement.

**'Prominent' attribution:** To that end we suggest the CRs mandate 'prominent' rather than 'sufficient' attribution and set out a **minimum standard** for both reader facing and back-end publisher facing aspects of attribution.

**Reader facing attribution in the body of an AIO or AIMode:**

- **The publisher's logo.** This is crucial for brand visibility and brand survival and should be included.
- **A link to the specific article** rather than just to the publisher's home page
- **A link at the point of consumption** rather than relegated below the AIO or behind a hidden menu.
- All the above should be included on an unobscured basis **inside the body of the text**, rather than below or to the side.

#### **Back-end publisher attribution (as above):**

- **Total Citations**
- **Accredited citations.**
- **Accredited citations as a % of grounding score**

#### Third-party crawling:

**Open-source dataset loophole:** We welcome the CMA's specific reference that Google should not circumvent publisher choices by acquiring opted-out content through other means (4.26ii) but would caution against the exemption to "open-source datasets....where these datasets have obtained content legally, given the open nature of that data source." In recent evidence to the House of Lords, <sup>5</sup>Roxanne Carter, Google's senior manager for government affairs, said *"When it comes to training AI models on **freely available content that is available on the open web**, we do not believe that we should license."*

**'Open web' or 'publicly available' does not equal free to use:** In accordance with its Terms of Service the Guardian makes its content available for very limited purposes to private, non-commercial readers - we have always been committed to making our journalism available to the widest possible audience; our content is not behind a paywall. We have also committed to making our journalism available for mission-driven motivations. . In addition, via internet protocols like robots.txt the Guardian has permitted crawling of its content for the single purpose of organic SERPs indexes to drive awareness, CTR and on-platform monetisation. None of the afore-mentioned undermines or conflicts with international copyright protections afforded to Guardian content. Accordingly, all other uses of Guardian content require a commercial licence from GNM. We consider any unauthorised use of our content (whether by third parties for AI training or other purposes) to be an infringement of the Guardian's copyright and breach of the Guardian's terms of service.

Increased crawling for wholesaling purposes (detailed in the recent 'State of the Bots' report by Tollbit<sup>6</sup>) is forcing publishers to tighten their IP protections. The Guardian is keen to ensure that our journalism remains open to the widest possible human audience. Exempting 'open-source datasets' is a potential loophole which could have the unintended consequence of disincentivising open access to quality content. To this end we would suggest wording along these lines:

<sup>5</sup> <https://committees.parliament.uk/oralevidence/17016/pdf/>

<sup>6</sup> [https://tollbit.com/state-of-the-bots/q3-q4-2025/?utm\\_medium=email&\\_hsenc=p2ANqtz--FZ\\_G-FwEWmY7Ljj96NjgCUIsQu-aiyrg8i6\\_u0x3OzQ4mi9utWHpysrB2i0Ove0tAhVQpRveWoJUwjhtwCopqFO\\_SSfotUyKEVi3KumkWE-XqTMA&\\_hsmi=402426378&utm\\_content=402426378&utm\\_source=hs\\_email](https://tollbit.com/state-of-the-bots/q3-q4-2025/?utm_medium=email&_hsenc=p2ANqtz--FZ_G-FwEWmY7Ljj96NjgCUIsQu-aiyrg8i6_u0x3OzQ4mi9utWHpysrB2i0Ove0tAhVQpRveWoJUwjhtwCopqFO_SSfotUyKEVi3KumkWE-XqTMA&_hsmi=402426378&utm_content=402426378&utm_source=hs_email)

"In relation to a particular publisher's content, and without affecting a publisher's remaining rights under UK copyright law, Google must be prohibited from using any third-party dataset, regardless of its origin or legal status, for any purposes (e.g. to train or ground its AI services) for which that publisher has exercised its opt-out rights with Google."

This reflects the position of UK copyright law that all rights are held by the copyright holder, except for those that it elects to license.

### **Mechanism for publishers to more easily communicate the reason for blocking content (6.4a)**

Our understanding is that under the CRs there would be two options for publishers to block content appearing in AIO - in advance or after it is published. Both would be done by exercising page level controls. The latter might be in response to issues with the content that mean the publisher wants it taken down, or concerns about how the AIO are presenting the content.

**Advance blocking (4.67):** There should not be any responsibility on the copyright holder to give a reason for excluding materials - opting out is the prerogative of a copyright holder.

**Take down post-publication:** Accuracy is a key principle of the Guardian's journalism and is covered in detail in our published [Editorial Code](#). Similarly news broadcasters are governed by the extensive and thorough Ofcom code. At present publishers' only recourse within AIO is a thumbs up/ thumbs down mechanism (4.64b) which is "then considered on an aggregate basis." The CMA's appraisal that this has 'limitations' downplays the damage that incorrect information about news events has on brand reputation, public safety and the wider information ecosystem.

**Removing content from an LLM:** While we welcome the option to opt out content at a page level we are concerned about the efficacy of the process. AI developers<sup>7</sup> and researchers<sup>8</sup> have frequently made the case that **it is not viable for LLMs to surgically delete ingested material** - clarification from Google that they have a technical solution for removing content would be helpful.

**Complaints mechanism:** If it is the case that once ingested, content remains in the model i.e. it cannot be unseen and so risks being reused by the model, then the CRs should specify a mechanism for raising complaints. This mechanism needs to be in real time, especially as there is a lack of English law precedent clearly assigning liability for inaccuracies in AI generated overviews. Without a timely complaints mechanism for publishers, users risk unknowingly being given inaccurate information.

---

<sup>7</sup> <https://cdn.arstechnica.net/wp-content/uploads/2024/11/NYT-v-OpenAI-Data-Deletion-Denial-11-22-2024.pdf>

<sup>8</sup> <https://arxiv.org/pdf/2407.06460>

## Accelerated timeframe:

In the time since the DMCC was passed, large-scale use of copyrighted content for generative AI has continued unchecked. The lag was extended by CR proposals not being published at the same time as designation - however Google will have had an indication of the direction of travel from the CMAs roadmap published in June 2025. Given that the "Google-Extended" toggle already exists, the proposed solution does not require them to build controls from scratch. GMG would support a shorter window of **3 months for implementation of the Publisher Control CRs is sufficient.**

## Recommended additions to the CRs

### **Explicit opt-in for onward sale or monetisation of publisher-derived data.**

There is a growing market<sup>9</sup> for scraping and indexing services which sell collated data for AI training and grounding to third-parties. Selling a search index to an LLM requires not just the index but the actual content - it represents an appropriation of value away from publishers to the search engine. The CRs should clearly stipulate that the sale, licensing, or commercial exploitation of such derivative representations requires explicit publisher consent.

**Clarity about dealing with a breach of compliance.** The CMA has left much of the detail of compliance in Google's hands. While GMG very much hopes that the CRs will achieve the stated aims, it would be helpful to robustly define the timeframe and process by which the CMA could move to a 'plan B' if needed.

### Splitting the Googlebot:

We appreciate that the CMA is under an obligation of proportionality, and is unlikely to reverse its position on splitting the crawler. However we wish to put forward the following arguments in the event that the CRs do not fulfil their aim.

**Real time visibility:** A split crawler could be monitored by publishers in real time and on an ongoing basis giving more visibility and certainty. Most publishers have already invested in bot monitoring systems so the workflows are already in place. A split crawler would allow publishers to independently monitor the impact of the CRs without waiting for periodic reports - reducing auditing costs and creating compliance efficiencies.

**Clarity of use:** Indexing-for-discovery and indexing-for-use are clearly separate concepts. The recently published Cloudflare bot principles<sup>10</sup> set out why responsible crawler practice requires single purpose bots. With the rise of stealth crawlers, there is a growing body of thought<sup>11</sup> that indexing should be split into retrieval-first indexing that can be used by LLMs and RAG (for AI and answers) and discovery-first indexing for [humans](#). We understand that the CMA has acknowledged Google's position not to split the crawler, however our position is that this overlooks a core problem and is not a forward looking solution to the issues presented by AI.

---

<sup>9</sup> <https://tollbit.com/state-of-the-bots/q3-q4-2025/>

<sup>10</sup> <https://blog.cloudflare.com/building-a-better-internet-with-responsible-ai-bot-principles/>

<sup>11</sup> <https://arxiv.org/abs/2512.00313?utm>

**Unverified cost argument:** Ruling out a split crawler partly on the basis of costs is fundamentally asymmetric as it ignores the costs that a less effective solution might have on the publishing industry. Having effective publisher controls should be seen simply as a cost of doing business. The £150m pa cited by Google as the cost for splitting the bot is small compared to its global revenue of over \$346 billion a year<sup>12</sup>. We are concerned by the lack of challenge by the CMA on this figure as it assumes:

(i) that the ‘ask’ is for them to duplicate all their infrastructure, when in reality, the ask is for publishers to have distinct controls for indexing for specific uses. **This is possible without duplicated infrastructure.**

(ii) that internet architecture, and in particular AI markets, must or will continue using crawler technology rather than **migrating to optimised datasets, API feeds and non-public datasets**. Crawlers have built-in inefficiencies e.g. crawler redundancy, dynamic content processing, content relevancy; they are also limited in their ability to sift content in a granular manner and differentiate between ownership profiles of crawled content (media businesses acquire a material percentage of their content, especially images, under licence from third party suppliers).

(iii) that a split Googlebot would impose significant incremental operating costs on third-parties including publishers. For GMG the **incremental cost of duplication of an extra crawler on our site would be minimal**. It also pre-supposes that all publishers would opt in to both crawlers. Furthermore Google already operate over 20 crawlers and fetchers<sup>13</sup> for a variety of purposes suggesting this is not a technical limitation.

**Competitive advantage:** In addition, allowing Google to spread the cost of operating its crawler across both search and AI gives it an **ongoing advantage over emerging LLM competitors** who do not have a lucrative search business against which to offset infrastructure or ongoing utility costs. This impacts end-user choice and service.

## FAIR RANKING CR

Is the FR scope correct (6.2)

We are pleased to see that the scope covers both placement in search and placement within AIO. We also welcome the specific reference to potential self-preferencing of Youtube videos above text content. However to ensure that the CRs fully achieve the stated aim we recommend the following:

(i) **The principles of Fair Ranking should also apply to the ranking of Google’s own search features relative to organic results.** If AIO are automatically prioritised over blue links and search is relegated below the fold then this entire process could be seen as arguing over scraps. The recent development of follow up questions in AIO automatically

---

<sup>12</sup>[https://www.statista.com/statistics/266206/googles-annual-global-revenue/?srsltid=AfmBOoqMXmenMF0yaHfhLhQ\\_4PUUIFNwVyNJpctO1GuXPmlSUg\\_WRaBX](https://www.statista.com/statistics/266206/googles-annual-global-revenue/?srsltid=AfmBOoqMXmenMF0yaHfhLhQ_4PUUIFNwVyNJpctO1GuXPmlSUg_WRaBX)

<sup>13</sup> <https://developers.google.com/crawling/docs/crawlers-fetchers/google-common-crawlers>

rolling into AI Mode (as detailed in [this Google blog](#)) demonstrate that without a fair playing field between search and overviews, the CR will be out of date before it is enacted - it can't wait for potential consideration in phase 2.

**(ii) The Fair Ranking CRs must be viewed in conjunction with the strength of the Publisher Control CRs.** In the case of news queries, the 'objective criteria' needed to judge whether an end user is better served by an AIO or a blue link may depend on the transparency and attribution of the sources in the overview. If they are poor - it impacts the user's ability to judge if a result is fair, accurate or true. In which case it becomes objectively preferable for the user to be served up the individual sources rather than an opaquely aggregated summary.

**(iii) Prioritise news carousels:** GMG's view is that when a news carousel is triggered, **users will have an improved experience if it is served above the AIO.** In 2024 Google said they would "not show AI Overviews for hard news topics, where freshness and factuality are important."<sup>14</sup> This analysis is backed up by the EBU's comprehensive report<sup>15</sup> on news answers from AI engines which found that:

- Almost half of all AI answers had at least one significant issue.
- A third of responses showed serious sourcing problems.
- A fifth contained major accuracy issues, such as hallucinated and/or outdated information.

To draw on Google's own language (Publisher Controls CR 4.64a) this is a way of 'optimising towards factuality... by using humans'. Users will still have the option to read the AIO, but they will have improved visibility of primary source journalism, with the ensuing verification processes and liability for the primary sources inherent in news organisations with an editorial code.

## Transparency

**Defining 'objective criteria':** We note the difficult trade-off to provide 'sufficient' information about ranking criteria, without revealing information that could then be gamed by spam or other bad faith actors. By not requiring Google to reveal the proxy signals it uses to rank search results, it makes non-discrimination almost impossible to enforce.

Transparency is key to the CMA's aim of mitigating Google's advantage afforded by their dominance in search and ensuring it does not give an uncompetitive advantage to their other services. While we appreciate that Google will not be able to disclose signals or weighting priorities, there is precedent for independent external auditing of the underpinning algorithms (eg casino games audited by the gambling commission or credit scores audited by the FCA).

**Retaliatory action:** While the CR explicitly forbids Google from taking "retaliatory action", proving that a ranking drop is the result of a commercial negotiation is notoriously difficult

---

<sup>14</sup> <https://blog.google/products-and-platforms/products/search/ai-overviews-update-may-2024/#:~:text=We%20updated%20our%20systems%20to,make%20Search%20better%20for%20ever,yone.>

<sup>15</sup> <https://www.ebu.ch/research/open/report/news-integrity-in-ai-assistants>

given the "black box" nature of algorithmic systems. We would **welcome additional information from the CMA about how they intend to police this.**

## Non-distortion and complaints

**Prescriptive complaints process:** Google's dominance in search means that manual actions to block or downgrade content can have an enormous impact on publishers. The current complaints process is opaque and slow. In the case of news publishers, a slow response may be akin to no response at all. The current language leaves the changes to Google's discretion. We suggest

(i) Insert the word 'timely' into paragraph 6 of the FR CRs i.e. "Google shall provide a clear, accessible and timely process for handling complaints from publishers ...."

(ii) an obligation on Google to explain why the content has been downgraded

(iii) an independent appeal process for issues that are not resolved in a clear, accessible and timely manner.

## USER CHOICE CR

### Frictionless opt out:

GMG strongly supports giving Google users a frictionless and transparent opt-out process from AI interfaces. Many Guardian readers want primary source journalism rather than AI summaries - as demonstrated by the percentage of views that come from branded queries i.e. using the word Guardian in the search bar when looking for information. To ensure user autonomy, Google should make it easy for users to signal their preference, and these preferences should be respected.

### Gemini inconsistency:

As referenced in previous GMG submissions we are disappointed that Gemini was removed from the scope of these CRs. But in light of this it seems inconsistent to define other AI assistants as equivalent search engines for the purpose of choice screens. Driving more users to unregulated platforms is likely to impact publisher direct traffic in a way that is at odds with the CRs aims.

## AUDITING MEASURES (for all CRs)

**Ongoing third-party auditing:** We agree with the CMA's provisional view that there is 'intrinsic value' in the Baseline Compliance Audit being conducted by an independent third party. But after the initial audit, the process moves to self-correction reports whereby Google investigates itself and then publishes reports on whether its policies are distorting markets. Although the CMA will have quarterly oversight of complaints, it risks leaving Google with leeway to determine what is market distortion v legitimate policy objectives. To improve

trust and confidence in the process, we strongly recommend that third-party assessment should be an ongoing requirement rather than a one off.

**Skilled person appointment:** Section 79 of the DMCC gives options for either the CMA or ‘the undertaking’ to nominate a ‘skilled person’ who can oversee the audits and reports. We strongly urge the CMA to adopt a 'Direct Appointment' model. The 'skilled person' must demonstrate a proven track record in auditing algorithmic transparency and data-scraping ethics, rather than just general compliance, to ensure that the audit results in actionable, granular data that can be verified by the publishing community.

We would also recommend that the following are clearly specified in the remit for the audit

**(i) Analysis of Internal data combining:** Even if publishers are able to set specific controls, Google could blend the data internally in opaque ways that are difficult to audit or reverse. This should be a specific area of examination for the ‘skilled person’.

**(ii) Level of zero-click searches:** Given that Google’s AI features aim to keep users within Google’s ecosystem, the level of zero-click searches should be specified as a key metric for the "skilled person" to audit.

**Publisher feedback:** The perspective of publishers is absent from the monitoring and compliance mechanism. Their views should be formally built into the CMAs oversight process.

## SUMMARY OF PROPOSALS

Area	Specific Measure / Recommendation
<b>Controls</b>	Mandate separate, defined opt-in controls for specific services: Search, Grounding, Fine-tuning, Training, and other named AI features .
	Require clear, detailed descriptions of the scope, purpose, and limitations of controls like Google Extended (GE) to avoid vague "catch-all" terms.
	Extend page-level and directory-level controls (using meta-data or tags) to all Google Extended features, including training and grounding.
	Standardise technical signals across features and implement a clear, simple opt-in process via Search Console (similar to cookie consent toggle).
	Disclose new AI products 30 days before launch to allow publishers to set controls in advance.
<b>Data</b>	Mandate feature-level reporting parity with Discover in Search Console, providing fully disaggregated data for every AI feature.
	Provide data allowing publishers to compare CTR for blue links in traditional search vs. those appearing beneath AI Overviews (AIO).

	Reintroduce URL referrers to provide visibility of user journeys on a per-article basis.
	Mandate granular metrics: Total Citations (by category/keyword), Accredited Citations (unique pages), and citations as a % of grounding score .
<b>Attribution</b>	Set a minimum standard: include the publisher's logo, links to specific articles (not just the homepage), and links inside the body of the AI text .
<b>Fair Ranking</b>	Apply fair ranking principles to Google's own features relative to organic results to prevent search from being relegated "below the fold".
	Prioritise news carousels by serving them above AI Overviews to optimize for factuality.
	Prescriptive complaints process that specifies timings and the reasons for manual actions
<b>Content Removal</b>	Clarify the technical solution for removing ingested content and provide a real-time complaints mechanism for inaccuracies.
<b>Third party data and uses</b>	Require Google to verify the provenance of third-party datasets and ensure they do not contain opted-out material.
	Require explicit publisher consent for the sale, licensing, or commercial exploitation of publisher-derived data/indices to third parties.
<b>Ongoing Auditing</b>	Mandate ongoing third-party auditing (not just a one-off) using a "Direct Appointment" model for a skilled person.
<b>User Choice</b>	Make it easy and clear for users to opt out of Gen Ai
<b>Implementation Timeline</b>	Accelerate the schedule to 3 months for implementation of publisher controls.