News/Media Alliance Comments Following Workshop on Promoting Competition in Artificial Intelligence

The News/Media Alliance (N/MA) respectfully submits these comments following the recent Department of Justice and Stanford University Workshop on Promoting Competition in Artificial Intelligence, held on May 30, 2024. We commend the Department for convening the workshop and for your attention to important questions posed by the proliferation of artificial intelligence (AI) systems and models, including the fundamental challenges they pose to news, magazine, and digital media publishers.

N/MA is a nonprofit organization representing the news, magazine, and digital media industries, and empowering members to succeed in today's fast-moving media environment. N/MA represents over 2,200 diverse publishers in the United States and internationally, ranging from the largest news and magazine publishers to hyperlocal newspapers, and from digital-only outlets to papers who have printed news since before the Constitutional Convention. Our members publish quality journalistic and creative content that covers natural disasters, conflict zones, school boards, city halls, townhalls, entertainment and the arts, food, wellness, health, technology, and other matters that keep our communities informed, engaged, and entertained.

The rise of generative AI models, systems, and applications presents a host of opportunities and challenges for publishers and creative industries as a whole. We applaud the Department of Justice for focusing on the importance of promoting competition in AI as we strongly believe that fair and efficient competition in the digital marketplace is vital for the sustainable development of generative AI systems, the future of copyright industries, and the sustainability of access to quality, diverse American journalism. We were encouraged by the inclusion of various creative industry voices in the workshop panel on creators' rights, and express broad support for the views discussed during the panel.

Our comment focuses on how generative AI developers misappropriate publisher content – undermining fair competition and risking our communities' access to high-quality information and media content – and how this misappropriation enables abusive practices, maintains existing monopolies, and threatens to create new ones. It also voices concerns over the potential transfer of publisher data to third parties by dominant search engines. Note that generative AI models and applications use publisher content in a fundamentally different way than other vital, non-generative use cases do. Accordingly, these comments are limited to issues

raised by generative AI and should not be considered as N/MA's views or positions on non-generative uses.

The Future of American Publishers Depends on Fair Competition between Rightsholders and AI Developers

Far too many media publishers are struggling. They face an existential threat to their ability to continue investing in high-quality content. In a properly functioning market economy, AI developers would acknowledge publishers' undisputable intellectual property rights in their content, and respect rules of fair competition by not free riding off the investments of publishers to build services which compete with the original sources of revenue. These rights would then facilitate fair exchanges of value between the owners and licensees. However, and with some notable exceptions, this is not yet occurring.

Instead, many AI developers misappropriate publisher content without compensation to train their generative AI models, ground those models, and retrieve real-time answers to user queries (sometimes known as Retrieval-Augmented Generation or "RAG"). AI model outputs often compete directly with the original works, providing content in sometimes repurposed forms, and diverting users from visiting the original source material, often failing to provide attribution to the source. In some instances, the outputs include seemingly realistic content that does not correspond to any real-world input – a "hallucination" – that is falsely attributed to a publisher, thereby misleading users as to the source of the information and damaging the publisher's reputation and brand. This misappropriation of publisher content is without authorization or compensation for such uses. Taking one egregious example, Mustafa Suleyman, CEO of Microsoft AI, recently stated in an interview:

> *I think that with respect to content that's already on the open web, the social contract of that content since the '90s has been that it is fair use. Anyone can copy it, recreate with it, reproduce with it. That has been "freeware," if you like, that's been the understanding*.[1]

The statement is obviously wrong on the law, and yet, is unfortunately not an isolated one. In March, Mira Murati, OpenAI's CTO told the Wall Street Journal that their new text-to-video product, Sora, was trained on "publicly available data and licensed data," and when asked whether this included videos from YouTube, stated: "You know, if they were publicly available — publicly available to use."[2]

[1] Sean Hollister, *Microsoft's AI Boss Thinks It's Perfectly Okay to Steal Content if It's on the Open Web*, THE VERGE (Jun. 28, 2024), https://www.theverge.com/2024/6/28/24188391/microsoft-ai-suleyman-social-contract-freeware.
[2] Maggie Harrison Dupré, *In Cringe Video, OpenAI CTO Says She Doesn't Know Where Sora's Training Data Came From*, YAHOO! NEWS (Mar. 15, 2024), https://au.news.yahoo.com/cringe-video-openai-cto-says-130138495.html.

The consequences of AI developers embracing and acting on this kind of thinking will be catastrophic to the free press. We are seeing blatant and systemic misappropriation of content and overt efforts to keep users on AI developers' own services that unfairly compete with publisher properties. Developers' actions to take and wield publishers' own content against them seriously compromises publishers' ability to monetize their content through advertising, subscriptions, and data. High-quality, accurate content is expensive to produce, and readers have long been willing to either subscribe or accept advertising and data sharing in exchange for low-cost access to publisher content. Without the ability to benefit from these longstanding revenue streams and with no new alternatives available, publishers across the country may be forced to close or seriously curtail operations which will cause the quality and quantity of information that supports our democracy to suffer.

Misappropriation of Protected Content Enables Market Abusive Practices That Help Enshrine Monopolies

While the misappropriation of content by AI developers of any size is worrying, we are especially concerned by the actions of the largest platforms, which are misappropriating publisher content to keep users and preserve their dominant positions in the digital marketplace. The problem is especially acute when the platform wields substantial control over publishers' ability to distribute and monetize their content on the web, as is the case with Google.

As discussed in our recent letter to the Department of Justice and the Federal Trade Commission,[3] Google's AI Overviews product (formerly, Search Generative Experience (SGE)), released nationwide on May 14, 2024, uses publisher content to generate comprehensive AI Overview answers to user queries – generating content derived from the content in its search index, including original and breaking news content. This Google-generated content is then displayed at the top of its search results, effectively allowing Google to "do the googling"[4] for its users and reducing demand for third-party websites.

According to some estimates, this integration of AI-enabled generative search and summarization may lead to 90 percent of search queries never leaving Google's services.[5] Uses like these – that keep users within the developers' walled gardens – combined with the systemic

[3] NEWS/MEDIA ALLIANCE, Letter to Assistant Attorney General Jonathan Kanter and Chair of the Federal Trade Commission Lina Khan (May 28, 2024), https://www.newsmediaalliance.org/wp-content/uploads/2024/05/2024-05-28-Coffey-Ltr-to-DOJ-and-FTC.pdf.

[4] Connie Guglielmo, *With AI, Google Wants to Do All 'the Googling for You.' Is That a Good Thing?*, CNET (May 20, 2024), https://www.cnet.com/tech/computing/with-ai-google-wants-to-do-all-the-googling-for-you-is-that-a-good-thing/.

[5] Omer Leibenzon, *Study Reveals: SGE Threatens 90% of Google's Top Searches*, ANGORA MEDIA (Mar. 31, 2024), https://www.angoramedia.com/blog/study-google-sge-impact-brands.

misappropriation of protected content for AI training purposes in the first place, can have profoundly anticompetitive effects in the digital marketplace. By inserting themselves between rightsholders and their audiences, AI developers disrupt and manipulate that relationship to redirect revenue flows, entrenching their market positions. In the case of Google, due to the lack of an effective way for publishers to opt out of or prevent unauthorized uses of their content for SGE without also opting out of search indexing, Google is effectively using its market position in search to force publishers to give over their content for free for generative AI purposes, protecting Google from market competition.

N/MA urges the government to safeguard against these imbalances and the misappropriation of protected content by AI developers. In the first instance, we reiterate our call for the Department and the FTC to use their authority under Section 2 of the Sherman Act and Section 5 of the FTC Act to stop Google's latest expansion of AI Overviews, investigate its monopolistic misappropriation of publishers' content, and take necessary action to end Google's abusive conduct.

Downstream Distribution of Publisher Content for AI Purposes by Large Online Platforms Raises Unique New Challenges

With the proliferation of Retrieval Augmented Generation technologies that provide AI users with real-time answers grounded in external, non-training data, search providers – namely Google and Microsoft – appear to offer APIs for external AI developers to gain access to their search index or aspects thereof.[6] The type and amount of information shared with third-party developers is unclear, although it seems that both companies provide at least titles, URLs, and snippets in response to API queries.[7] Providing information from these indexes to third-party developers has the potential to compound competition challenges.

---

[6] *See, e.g.*, OpenAI, How do I use ChatGPT Browse with Bing to search the web?, https://help.openai.com/en/articles/8077698-how-do-i-use-chatgpt-browse-with-bing-to-search-the-web ("When the Browse feature is used, ChatGPT formulates a keyword search based on your prompt and submits search prompt to a headless version of the Bing search engine to retrieve relevant results.") (last accessed Jul. 15, 2024); Google, *Ground Responses for Gemini Models*, GOOGLE CLOUD, https://cloud.google.com/vertex-ai/generative-ai/docs/multimodal/ground-gemini (last accessed Jul. 5, 2024); Google, *Grounded Generation API Pricing*, GOOGLE CLOUD, https://cloud.google.com/generative-ai-app-builder/pricing#grounded_generation_api_pricing:~:text=Media%20Recommendations%20errors.-,Grounded%20Generation%20API%20pricing,-The%20Grounded%20Generation (last accessed Jul. 5, 2024); Google, *Custom Search JSON API*, PROGRAMMABLE SEARCH ENGINE TUTORIAL, https://developers.google.com/custom-search/v1/overview (last accessed Jul. 5, 2024); Microsoft, *Use and Display Requirements of Bing Search APIs, with Your LLM*, BING WEB SEARCH API DOCUMENTATION, https://learn.microsoft.com/en-us/bing/search-apis/bing-web-search/use-display-requirements-llm (last accessed Jul. 5, 2024).

[7] Google, *Using REST to Invoke the API*, PROGRAMMABLE SEARCH ENGINE TUTORIAL, https://developers.google.com/custom-search/v1/using_rest#response_data (last accessed Jul. 5, 2024) ("The search results include the URL, title and text snippets that describe the result. In addition, they can contain rich snippet information, if applicable."); Microsoft, *Use and Display Requirements of Bing Search APIs, with Your LLM*, BING WEB SEARCH API DOCUMENTATION, https://learn.microsoft.com/en-us/bing/search-apis/bing-web-search/use-display-requirements-llm (last accessed

The Google and Bing search indexes were originally created for use by Google and Microsoft to power their core search engines. To be indexed by these search engines and appear in search results, publishers have, reciprocally, permitted their sites to be crawled for these specific purposes – and only for these purposes. Publishers agree to be indexed because search is a vital distribution channel that directs users to the publishers' original content, thus increasing traffic and revenue. But search distribution is fundamentally different from this new practice where publisher content is used for generative AI that diverts rather than augments viewership.

Competition authorities should be wary in general of Google and Microsoft using their leverage over publishers to force publishers to give up their content for this alternative use. These large technology platforms are effectively using their market power to monetize (for their own benefit and not for the benefit of publishers) content that does not belong to them.[8] In doing so, they are also making it easier for other developers to gain unauthorized access to publisher content. This has the potential to unfairly benefit AI developers who would otherwise have to acquire and negotiate access to this content directly with the rightsholders.

The Department of Justice should consider requesting more information about how the dominant search platforms sell or make available their search index materials to third-party developers.

The Government Can Help Correct Market Imbalances and Facilitate a More Competitive AI Ecosystem

To protect and encourage the development of an innovative, fair, and vibrant digital marketplace and AI ecosystem, the federal government should proactively address and enforce against the worst excesses of market dominance and abusive behavior by generative AI developers.

The Department of Justice should investigate Google's anticompetitive actions in relation to its AI Overviews offering, as well as proactively monitor, investigate, and enforce against other harmful market practices and misappropriation of publishers' content by AI developers. We commend the Department for its ongoing efforts to address the anticompetitive practices by

---

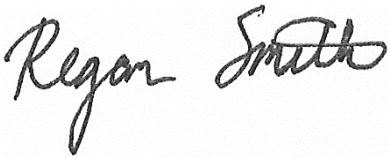
Jul. 5, 2024) (defining "web results" as "[t]he title, URL and snippet for the top ten webpage results returned from the Bing Web Search API.").

[8] Both companies seem to charge for the commercial uses of their APIs. *See, e.g.*, Google, *Grounded Generation API Pricing*, GOOGLE CLOUD, https://cloud.google.com/generative-ai-app-builder/pricing#grounded_generation_api_pricing:~:text=Media%20Recommendations%20errors.-,Grounded%20Generation%20API%20pricing,-The%20Grounded%20Generation (last accessed Jul. 5, 2024); Google, Custom Search JSON API, Programmable Search Engine Tutorial, https://developers.google.com/custom-search/v1/overview (last accessed Jul. 5, 2024) ("Custom Search JSON API provides 100 search queries per day for free. If you need more, you may sign up for billing in the API Console. Additional requests cost $5 per 1000 queries, up to 10k queries per day."); Microsoft, *Bing Search API Pricing*, BING SEARCH API, https://www.microsoft.com/en-us/bing/apis/pricing (last accessed Jul. 5, 2024).

the dominant online platforms, including through the existing Google enforcement actions. In addition, the government should also support efforts to encourage marketplace licensing, including on an aggregate basis, such as voluntary collective licensing initiatives.

Thank you for your attention to this important issue. We appreciate the opportunity to provide these comments and stand ready to answer any questions you may have.

Sincerely,



Danielle Coffey
President & CEO
News/Media Alliance

Regan Smith
Senior Vice President & General Counsel
News/Media Alliance