

**Senator Josh Hawley**  
**Questions for the Record**

**Danielle Coffey**  
**President and CEO, News/Media Alliance**

**1. Are Google, Meta, and other social media responsible for a great deal of your members' online traffic?**

Meta and Google control a large majority of the digital marketplace, capturing more than [70% of all digital advertising revenue](#), leaving publishers with no choice but to play by the tech platforms' rules if they want the referral traffic, which many of them rely on.

**Google:**

Google has built a walled garden in which it tries to keep users from leaving the platform by providing increasingly detailed snippets taken directly from news and other websites that it displays inside search results. In this way, users never have to click through to the source to read more, a phenomenon known as zero-click search.

A study by [Sparktoro](#) found that from January to December 2020, 64.82% of searches on Google (desktop and mobile combined) ended in the search results without clicking to another web property. Looking at mobile only, the percentage increases to more than three-quarters (77.22%) of mobile searches that end without a click.

According to a [white paper](#) published by the News/Media Alliance on “How Google Abuses Its Position as a Market Dominant Platform to Strong-Arm News Publishers and Hurt Journalism,” “Google’s use of news publishers’ content does send substantial traffic to news publishers, but Google is not fairly or appropriately compensating news publishers for the value of their material, or properly treating the news industry as an important strategic partner. Instead, Google has misused its position as the dominant online platform to reap the benefits of the news media’s substantial investments in reporting without paying a license fee. Further, it has misused its monopoly power to remove the ability for news publishers to have adequate control of the use of their content – often using its market power to force publishers into granting Google the apparent right to make vast and unknown uses of their intellectual property far into the future, or other problematic conditions, if they wish to be included on basic Google services.”

The [New York Times](#) reported that publishers are increasing their reliance on Google as a result of Facebook’s algorithm changes (though some publishers have seen decreased traffic from Google as well).

**Meta/Facebook:**

According to reports from publishers and data from [Echobox](#), traffic from Facebook has been on the decline since peaking in summer 2022, but “An apparent change to Facebook’s algorithm in

May [2023] caused a dramatic drop in traffic to news and media websites.” Traffic from Facebook fell 50 percent in the course of a year.

Looking at traffic globally, social media referrals decreased from 9.3% to 5% since 2020. “Both Facebook and X (Twitter) have been deemphasizing the display of news links that might cause people to click away from their platforms,” according to [Similarweb](#). The trend can be seen with [all news publishers](#), including [legacy and digital-first](#) news publishers.

In 2023, Meta decided to remove news from Facebook in Canada following passage of a bill in that country that would require the tech platforms to collectively negotiate with news publishers for compensation for the use of their online content because they didn’t like it and didn’t want to pay publishers for their content. The move coincided with dangerous wildfires in the country. In some less populated areas of the country, users rely on updates on Facebook to stay informed, and without it, their safety was put at risk. However, Meta/Facebook was more interested in making a point and protecting their bottom line than the safety of their users. Contradictory to what Meta claimed, that news wasn’t an important component of its platform, Canadians said they came to Facebook for news. Similarly, when Australia passed the News Media Bargaining Code in 2021, [Facebook removed news](#) from the platform, including public health and safety organizations, in protest. After backlash in the country, the platform reinstated news the following week.

## **2. Do you believe that Google, Meta, and other big tech have an unfair advantage when it comes to bargaining for use of your members’ content?**

Google, Meta, and Big Tech have consistently shown that their market dominance gives them an advantage in bargaining and allows them to use our members’ content without authorization or fair compensation. And when traffic does it make it to our members’ sites, for each dollar spent on digital advertising these platforms take as much as 70 percent of the revenue, leaving publishers with only 30 percent.

Right now, there is no equitable bargaining— Big Tech is using publishers’ content without appropriate compensation, and yet publishers have no choice or reasonable alternative than to accept monopolistic terms from Google and Meta. Individual news publishers cannot effectively challenge the terms offered by these Big Tech companies, even though the terms do not recognize and compensate news publishers for the [value](#) they bring to their platforms in the form of traffic, data, and advertising revenue. As addressed in greater detail below, the emergence of generative artificial intelligence (GAI) only exacerbates the issue.

The market imbalance is a leading cause of the demise of local news across the country, as highlighted in a [report](#) released by the Medill School of Journalism at Northwestern University in November that found that the rate of newspaper closures across the country has accelerated, now at 2.5 closures per week (with more than 130 confirmed newspaper closings or mergers in the last year), resulting in the expansion of news deserts, in which communities lack a source of local news.

There must be a fair exchange for the significant value news publishers' content provides to Google, Meta, and other Big Tech companies so that publishers across the country have the revenue to invest in high-quality, trusted journalism. The bipartisan, Senate Judiciary Committee-reported Journalism Competition and Preservation Act (JCPA) is reasonably and specifically tailored to ensure news publishers can negotiate fair compensation with Big Tech to ultimately help protect and improve the quality and accessibility of reporting and ensure local journalists can continue their critical work.

**3. To your knowledge, have any of these companies approached your members to ask for permission to train their AI models on your content? Or did they simply do it without asking?**

As a general rule, AI developers do not ask for permission before they use our members' content to train their AI models. In some cases, content was scraped directly from N/MA members, and in other cases, companies made use of third party datasets containing N/MA member content. We are aware of emerging reports of licensing deals that address training for GAI models, which is an encouraging sign. We also know it is possible for developers to obtain permission because in many cases they have existing licensing agreements in place with our members for other uses, or have licensed material for AI training for other types of content.

**4. Do you believe the AI models developed by these companies will increase their market power?**

Yes, advances in GAI pose several risks to competition. In addition to increasing their dominance in online advertising as described below in Response 7, GAI could expand Big Tech's general market power as well as their dominance in news distribution particularly. Developers train their models using high-quality news content without consent or compensation and then generate outputs that include summaries, excerpts, and even full, verbatim copies of articles written and fact-checked by human journalists. These outputs usurp the creative expression and significant investment of news publishers and journalists while diverting news publishers' audiences to their platforms. By taking the work of local news outlets and other trusted news sources and integrating GAI into search and social media products, Big Tech will further draw consumers away from news publishers and into their walled gardens, increasing their market power and entrenching their position as gatekeepers to news and information online.

Big Tech companies may also use their dominant positions in data and computing markets today to secure their future dominance in AI. The Federal Trade Commission (FTC) [raised](#) concerns to the U.S. Copyright Office in comments that warn that these companies could use their data and computing resources to "unlawfully entrench their market positions in AI and related markets." The FTC stated:

The rising importance of AI to the economy may further lock in the market dominance of large incumbent technology firms. These powerful, vertically integrated incumbents control many of the inputs necessary for the effective development and deployment of AI tools, including cloud-based or local computing power and access to large stores of training data. These dominant technology companies may have the incentive to use their

control over these inputs to unlawfully entrench their market positions in AI and related markets, including digital content markets. In addition, AI tools can be used to facilitate collusive behavior that unfairly inflates prices, precisely target price discrimination, or otherwise manipulate outputs.

Just last month, the FTC [issued](#) orders requiring the disclosure of information related to investments and partnerships between GAI companies and cloud service providers and held a Tech Summit to explore the impact on competition of concentrated market power in markets for cloud infrastructure and graphics processing units (GPUs) used for AI processing. Congress should take action to ensure the tech platforms that have monopolized online advertising, search, news distribution, cloud computing, and semiconductor manufacturing, are not able to use GAI to further their dominance in the AI era.

**5. From your experience with these companies in the past, do you think the increase in their market power will be good or bad for you members?**

Any further increase in the market power of these giant tech companies will only exacerbate the harms that they are already inflicting on our members. As a result of Google's monopoly position in the online search market, our members are effectively forced to make their content available to Google's web crawlers, unless they want to be virtually invisible on the internet. An increase in market power will be bad for our members, who will lose even more bargaining power. For example, because publishers are forced to make their content available for search, Google may try to "ground" the AI systems that they are developing with fresh news content, developing service that compete with the very publishers that are feeding their models, further diverting traffic from the publishers themselves. This will decrease revenue for publishers, leading to further layoffs of journalists and more closures of at-risk publications.

**6. What about their readers and viewers?**

As increased market power for tech giants shrinks the news industry, news readers will suffer. They will have fewer sources for reliable, vetted information. Communities, especially smaller ones, will see less coverage of public affairs—or no coverage of local affairs at all. In communities that lose their local news outlet, municipal borrowing costs will increase. A [study led by Notre Dame](#) found that "following a newspaper closure, municipal borrowing costs increase by 5 to 11 basis points." Even pollution is likely to increase—a [study in the Journal of Environmental Economics and Management](#) found that when newspapers reported on pollution from consumer manufacturing plants, their toxic emissions were reduced by 29%.

As reliable news sources dwindle, more readers will get their news from unreliable sources who do not conduct the information vetting that professional journalists do, and often are intentionally promoting a biased viewpoint. This exacerbates the risk of intentional disinformation campaigns. And the news information provided to news readers by the GAI products themselves, which are already [disturbingly prone to false "hallucinations"](#), will become less and less reliable as the products are increasingly fed on unreliable inputs as reliable news outlets diminish and disappear, further depriving readers of the accurate information necessary for an effectively functioning democracy.

**7. Do you believe these companies could use generative AI technology to divert advertising dollars to themselves?**

Yes. In January 2023, the U.S. Department of Justice (DOJ) filed suit against Google alleging that it monopolizes key digital advertising technologies that website publishers, including news publishers, rely on. DOJ [alleged](#) that Google's DoubleClick for Publishers ad server has more than 90 percent market share and Google Ads advertiser network has more than 80 percent market share. According to DOJ, "Google pockets on average more than 30 percent of the advertising dollars that flow through its digital advertising technology products; for some transactions and for certain publishers and advertisers, it takes far more."

With Meta and Google already dominating the online advertising and news distribution markets, news outlets across the country are facing substantial revenue shortfalls. As business models have changed over recent decades, revenue from consumer subscriptions, licensing, and advertising is now fundamentally threatened by the rise of GAI. While today, news publications rely on Google's advertising infrastructure for revenue and Google uses robust snippets and summaries to keep viewers on their platform and away from publishers, if allowed to continue without obtaining required permissions, GAI could fully remove publishers from the revenue stream by unlawfully copying and regurgitating news content through their GAI products and services.

**8. Please explain your view on why using published content for training AI models is not (or should not be) fair use.**

In October 2023, News / Media Alliance published a white paper titled "How the Pervasive Copying of Expressive Works to Train and Fuel Generative Artificial Intelligence Systems is Copyright Infringement and Not a Fair Use." Our White Paper attaches a forensic analysis showing the pervasive use of media publisher content for training AI models. The White Paper, along with supplementary commentary relevant to fair use submitted in connection to an open study by the Copyright Office, is available [here](#).

**9. OpenAI provides outlets the ability to request that their content not be used to train AI models. Do other companies provide your members with the same option?**

Some AI developers, including OpenAI and some others, have provided publishers with opt-out measures, including honoring the robots.txt exclusion protocol on a go-forward basis, and most publishers also prohibit AI scraping through their terms of service. Increasingly many companies have opted out or blocked AI web crawlers—over the course of three weeks in late September at least 250 top websites blocked OpenAI's GPTBot while 14 percent of the 1,000 most popular websites block Common Crawl's CCBot.

Overall, opt-out measures, including technical measures such as robots.txt, are blunt and flawed instruments when it comes to protecting publishers from infringement in practice. Without cooperation from GAI developers, there is no easy, standardized way to block scraping for AI purposes. And blocking for AI training can often have the undesirable effect of also blocking

crawling for search and other desirable, mutually beneficial uses. There is also no requirement for developers to comply with the opt-out signal or for scrapers to accurately identify themselves, allowing bad actors to continue scraping publisher content without authorization. Further, and more fundamentally, publishers should not have to affirmatively opt out from generative AI uses to prevent the commercial consumption of their protected material—it is antithetical to the guiding principles of U.S. copyright law and the exclusive rights afforded to rightsholders. Such opt-out solutions are also “too little, too late,” considering the vast scraping and copying of publisher content that has already taken place to bring GAI models to the point of commerciality.

**10. OpenAI claims that regurgitation of published content is a “rare bug.” Have any of your members’ content ever been, to your knowledge, regurgitated in such a “rare” occurrence?**

Unfortunately, numerous reports show that the ability to reproduce publisher content verbatim is a feature, not a bug, of Large Language Models (LLMs) like OpenAI. Our members’ content is often regurgitated, paraphrased, or summarized in a manner that demonstrates that models have trained and retained publisher content. For example, The New York Times, an N/MA member, recently filed a lawsuit against OpenAI that includes an exhibit listing one hundred examples of verbatim regurgitation. Other reports, including [N/MA’s GAI White Paper](#), document similar phenomena.

Academic research disproves the idea that the models do not memorize expression. One recent [paper](#) noted, “by querying the model, we can actually extract some of the exact data it was trained on.” That study, conducted by researchers from Google DeepMind, the University of Washington, Cornell, Carnegie Mellon University, the University of California Berkeley, and ETH Zurich, documented bypassing of “alignment” rules to extract verbatim training data from open source, semi-open, and closed models. [Reporting](#) from *404 Media* on the study noted, “some of the specific content published by these researchers is scraped directly from CNN, Goodreads, WordPress blogs, on fandom wikis, and which contain verbatim passages from Terms of Service agreements, Stack Overflow source code, copyrighted legal disclaimers, Wikipedia pages, a casino wholesaling website, news blogs, and random internet comments.” While AI developers may attempt to mitigate after the fact by throwing up guardrails, this does not change the fact that the models themselves seem to have retained verbatim expression, and the guardrails are often faulty.

**11. Do you anticipate that generative AI will be used to aid the production of news media in the near future?**

**12. Is AI already being used to aid the production of news media? In what capacity?**

AI tools have been used for years by journalists. Publications have been cautious around the use of AI in newsrooms, building in ethical safeguards and best practices to ensure that journalists use AI-generated content in a transparent, accurate, fair, and accountable way. A [Reuters Institute survey of newsrooms](#) found that 68% of newsrooms have policies for the responsible use of AI released or in progress. 40% of newsrooms are designating a person to lead all



editorial aspects of GAI, and 38% are developing training and education programs for employees of the dangers associated with GAI.

Newsroom AI codes of conduct have many common elements, including:

- Disclosure in the article that an AI tool was used to assist the journalist.
- AI content must be verified with an “on the record” source.
- Prohibiting AI-generated photorealistic images in coverage.
- Outright prohibition of AI-generated content altogether.
- Prohibiting AI tools that were trained on copyrighted works of others.
- One large company has an AI Council and approves uses of AI tools in newsrooms. A journalist needs an executive sponsor, editor approval, and AI Council approval.

Humans are financially valuable to publishers. Publishers will generate more revenue by keeping journalists through higher advertising rates when there’s human involvement in the creation of the content. For example, [overtone.AI](#) checks text for journalistic signals that demonstrate human involvement.

Generally speaking the handful of articles that have gained notoriety for having been generated by AI is a grain of sand compared to the massive scale substitution for news by GAI.

One researcher who studied newsroom AI policies [found](#): “One thing that was remarkable to me is that the way in which organizations dealt with AI at this stage did exhibit a very strong sense of conserving journalistic values. Many organizations were really concerned about not losing their credibility, not losing their audience, not trying to give away what makes journalism stand out — especially in a world where misinformation is around in a much larger scale than ever before.”

### **13. Do you anticipate that generative AI models could replace current employees at news media companies?**

Yes, if GAI developers continue not to compensate news media publishers for the use of our content, we anticipate that they will leverage their substantial-to-overwhelming market power to deploy products that will compete with and displace publishers, leading to a concomitant reduction in journalists employed at news media companies.