

**Statement of Danielle Coffey
President & CEO, News/Media Alliance**

**U.S. Senate Committee on the Judiciary
Subcommittee on Privacy, Technology, and the Law
January 10, 2024
Oversight of A.I.: The Future of Journalism**

Chairman Blumenthal, Ranking Member Hawley, and Members of the Subcommittee, thank you for inviting me to testify here today on “Oversight of A.I.: The Future of Journalism.” My name is Danielle Coffey, and I am President and CEO of News/Media Alliance (N/MA), representing over 2,200 news, magazine, and digital media publishers, 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.

My members produce quality journalistic and creative content that seeks to inform, educate, and connect with readers and enrich their daily lives. Our publications adhere to principles and processes that support verification, accuracy, and fidelity to facts. We cover natural disasters, conflict zones, school boards, city halls, townhalls, entertainment and the arts, and other matters of public interest to local, national, and international communities. We abide by standards and codes of conduct and provide readers a voice through correction policies to ensure accuracy in reporting.

Unfortunately, I cannot say the same for most of what is disseminated and proliferated across the rest of the Internet. Without proper safeguards, we cannot rely on a common set of facts that promote healthy public discourse. Without quality reporting, we cannot have an informed electorate and functional society. This is particularly concerning as we start an election year where readers will rely on accuracy of information to make important decisions. With an already flooded market of falsified facts, Generative Artificial Intelligence (GAI) could create an even greater risk to the information ecosystem with inaccuracies and hallucinations if not curbed with quality content that comes from news publications.

The risk of low-quality GAI content dominating the Internet is amplified by the drastic economic decline of news publications over the past two decades. Studies [show](#) that more than two newspapers are closing each week, with one third of the countries’ publications set to close before 2025. These are local, community newspapers that cover neighborhood schools, football games, business openings, and make up the fabric of our country. Although local news outlets are fighting for their lives, people are interested in the news more than ever. While there has been a decline in revenue of 56% since 2014, over that same period of time, traffic to the top 46 news sites is up 43%. Why this increased consumption and output of quality journalism, yet the steep decline of financial intake? Because the dominant distributors of news content—also the world’s largest Internet companies—are scraping publications’ websites and selling portions for engagement and personal information to target users with advertising.

Over the last several years, there have been countless studies, investigations, and litigation by the Department of Justice in the past two Administrations that have found anticompetitive conduct by the monopoly distributors of news content. There is a direct nexus between this anticompetitive conduct and the steep decline of financial revenue returned to news publications. This has led to a dramatic shortfall of the revenue that is needed to fund costly expenses associated with newsgathering and employing journalists.

These dominant platforms have cornered the market on user data and targeted advertising through their walled gardens that they have created to keep users engaged, in part through news content my members create. Though fewer users are clicking through, and when users do click through, Big Tech monopolists extract unfair fees. This Committee and the last two Administrations have been actively scrutinizing the monopoly power in the ad tech ecosystem that results in up to 70% of advertising payments going to the dominant platforms. The extortion and arbitrage are indisputable. Because more than 70% of our referral audience comes from just two platforms, we are forced to rely on them, and accept these terms. Therefore, it is no surprise that publishers are not seeing revenue despite massive demand for our content. We are deeply appreciative of Senators Klobuchar and Kennedy for their leadership on the *Journalism Competition and Preservation Act*, which addresses this marketplace imbalance, and we thank this committee for passing this legislation that is so vital to the news industry last July.

However, as dire as the current dynamic is, this marketplace imbalance will only be increased by GAI. GAI developers crawl websites and reach behind paywalls to train their models. An analysis that the N/MA commissioned, as well as complaints pending before the courts, demonstrate the significant ingestion of quality news content in AI training models. Adding insult to injury, GAI “output” results to user inquiries often contain summaries, excerpts, and even full verbatim copies of articles written and fact-checked by human journalists. These outputs compete in the same market, with the same audience, serving the same purpose as the original articles that feed the algorithms in the first place. GAI is an exacerbation of an existing problem where revenue cannot be generated by, but in fact is diverted from, those who create the original work.

In the long term, protecting creators and rights holders from the unauthorized use of their works for training GAI will help both technology companies and news publishers provide better products and services to consumers. GAI models and products will not be sustainable if they eviscerate the quality content that they feed upon. Copyright-protected, expressive works have been taken without authorization and without consent, and used repeatedly in model training, processing, and display. Because these uses go far beyond the guardrails set by courts, now openly substituting for copyrighted content and usurping licensing markets, this should not be considered fair use under current copyright law.

To be clear, news publishers are not opposed to generative AI technologies—we want to help developers realize their potential in a responsible way. N/MA’s members are by and large willing

to come to the table and discuss reasonable licensing solutions to facilitate reliable, updated access to trustworthy and authoritative content. A constructive solution will benefit all interested parties and society at large and avoid protracted uncertainty. Some GAI developers are good actors and seek partnerships and licensing agreements with news publications, and we applaud their efforts. However, there is more work to be done.

With dominant search engines that combine AI datasets with real-time search results, a process known as “retrieval augmented generation” or “grounding,” antitrust laws should protect users from these anticompetitive behaviors. The practice of reliance on one dominant outlet, in this case search, gives publications little choice as to whether they would permit inclusion and acquiescence to this business arrangement. This is what’s known as a Hobson’s choice, or in antitrust terms, tying arrangements. Such an arrangement would provide little-to-no return to a publication where there is even less likelihood the user will click through.

Unless some marketplace corrections are immediately seen, we call on Congress to step in and investigate these practices. We also offer the following suggestions for Congress to ensure GAI advances in a sustainable manner:

Transparency, explainability, and traceability: *Congress should support legislation that requires the recordkeeping and disclosure of unauthorized training uses of material that is protected by copyright, by technical protection measures, or governed by contractual terms prohibiting scraping to disclose the use and weighting of specific usage of third-party content.* Substantial transparency measures must develop around the use of copyrighted materials in GAI technologies. The public should be able to know what AI models were trained on, and make the evaluations needed to select more ethically sourced or reliable models if they choose. And publishers have a right to know who copied their content and what they are using it for. The incentives to avoid disclosure are too strong to bet on a self-regulatory solution. Obligations should be tailored to scrapers, developers, and those who configure foundational models into customized applications.

GAI developers should be required to provide explanations detailing how their models produce outputs and provide links to materials cited in summaries. While there is value in international harmonization, and addressing other data-related concerns together, any outcome should achieve the core objective of providing sufficient transparency into the ingestion and use of copyrighted materials to allow rights holders to sufficiently analyze such models.

Copyright infringement and other harmful usage: *The unauthorized copying of publisher content to train and fuel commercial systems that produce substitutional output must be recognized as infringing.* Policymakers should push industry to acknowledge that the rampant copying of expressive media content to train LLMs that then compete with that content violates publishers’ exclusive rights and unfairly usurps their markets. N/MA believes that existing law establishes that this systematic and

competitive conduct is infringing. And we are heartened by emerging signs of licensed GAI uses for other creative works, like music and images. But wider recognition that media content is not free for the taking is critical to foster meaningful negotiations between GAI developers and publishers.

Congress should also consider legislation clarifying that GAI developers and deployers are responsible for the design of their technology, and do not qualify for safe harbors set up decades ago to encourage a nascent tech industry to host content without having to actively monitor and remove harmful material created wholly by third party persons. GAI systems are not passive hosts or conduits to content created or used by others—these systems actively use and generate content themselves. GAI systems and those who develop them should be held responsible and accountable, just like any other business.

Licensing and competition: *Policymakers should encourage market-based licensing solutions, and honor established law and policy that discourages government regulation of licensing markets as a first resort.* For some markets, this can include voluntary collective licensing as already permitted under law. Government should prevent developers from conditioning or modifying the provision of other services, such as advertising or search ranking, on (a) a content owner or site operator making available content for training or (b) a content owner or site operator permitting use of such content or site in search or other services or imposing reasonable terms and conditions.

Responsible design and accountability: *Congress should ensure generative AI development is designed to be responsible, rather than mitigating after the fact. It would be useful to make explicit that the immunities provided by Section 230 of the Communications Decency Act do not apply to AI-generated content.* Publisher experience in other contexts is that lowered standards for liability reduces the incentives for platforms to negotiate for uses of quality content, increases harm to the public, and places media publishers at a competitive disadvantage. Early GAI trends reveal so-called “overfitting” or “unintentional” harms to be all-too common. Developers should be incentivized to incorporate safety by design principles and maintain programs to prevent dangerous outcomes of AI-generated content, such as illegal content and other serious online harms.

Enforcement and anti-piracy: *Congress should ensure adequate tools to prevent unwanted scraping and “laundering” of copyrighted content.* Web scrapers must respect and follow terms of use and abide by automated flags that signal that online content be limited to specified uses. For example, tools could flag users’ desire to block crawling for training while permitting beneficial uses, or for training to be limited to particular uses or users. Known pirate sites should be off-limits for AI training purposes, even if those site owners would allow data scraping, and enforcement efforts of the U.S. Department of Justice, the Office of the Intellectual Property Enforcement Coordinator (IPEC), and other parts of government should be empowered and funded.

The Fourth Estate has served a valuable role in this country for centuries, calling on governments and civic leaders to act responsibly in their positions of power. Local news, especially, has uncovered and reported on events around the country that keep readers informed, educated, and engaged in their communities. We simply cannot let the free press be disregarded at the expense of new and exciting technologies. Both can exist in harmony, and both can thrive. We must ensure that for the future of our country and the future of our society.

Thank you for the opportunity to participate in this discussion. We look forward to working with Congress, the Administration, the States, and our counterparts around the world as this important discussion moves forward.