Dear Director Perlmutter,

The Open Markets Institute (OMI) welcomes the opportunity to comment on how the development of Generative AI (GAI) affects copyright. Decisions about how to apply copyright law to both the inputs and outputs of GAI will have profound economic and social consequences, making it imperative that the original public interest orientation of copyright law remain the North Star as the Copyright Office develops its guidance. The stakes are particularly high because the development of GAI is dominated by Big Tech corporations that already hold significant power over our information and communications channels. To avoid replicating the harms of surveillance capitalism and social media, the Copyright Office should therefore enforce copyright on input data and restrict granting copyrights to GAI-generated material and content.

In addressing several of the inquiry questions, this comment applies a competition policy lens to assess how copyright laws would apply to AI models given the dominance of a handful of companies in the AI space, from data to computational power to cloud services. This perspective is essential to ensuring that copyright policy is appropriately aligned with its public interest mandate. Open Markets is run largely by journalists and has long focused on the intersection of journalism and concentrated corporate power and control, including through our Center for Journalism and Liberty program. In this comment, we focus especially closely on journalism and the news media industry. We believe that in the absence of clear copyright guidance, GAI can exacerbate existing harms by Big Tech monopolies over the digital infrastructure upon which news production and distribution relies. Copyright policy cannot be properly constructed without considering its effects on market competition.

**Copyright’s Public Interest Purpose and the Present Landscape of Power in GAI**

The Copyright Clause of the U.S. Constitution states that the purpose of copyright law is to ensure that copyright protections advance the public interest. Therefore, the primary beneficiary should be the public. Amidst the discussion of how copyright laws and GAI intersect, some critics have argued that applying copyright law to this new market either “misses the mark” on fair use, stifles innovation, or threatens the openness and accessibility of the web. We disagree. Copyright is a critical part of structuring the market and will shape competition in ways that will ensure that sectors like journalism and the creative industries remain viable.
Promoting the public interest requires balancing multiple stakeholders and underscores the inseparable relationship between copyright law and competition policy. On one side, copyright establishes property ownership over the creative arts and published material. On the other, competition policy structures how the public can use those copyrights as part of a business venture using lawful methods of competition. In the case of GAI, its rollout can create billions of dollars in potential value or lead to billions in legal damages. Thus, defining whether copyright law applies to the GAI models, including the data inputs used to create them and the outputs these models generate, will fundamentally shape market competition.

We urge the Copyright Office to center the public interest mandate of the Copyright Clause by considering 1) how copyright law encourages and promotes human creativity and rewards labor through the protection it offers to individual journalists, writers, artists, and other creators; and 2) how such individuals can obtain financial rewards and recognition for their intellectual property used in GAI.

GAI is different from other recent technological shifts that have adversely affected the public, both in terms of the rate of adoption and disruption. Consider the difference in the speed of adoption. Google and Meta (formerly Facebook) took over a decade to attract hundreds of millions of users, redefine advertising and markets, reshape journalism, and become near-permanent fixtures in our daily lives. To reach 100 million users, it took ChatGPT only two months; largely because it was able to exploit the size and power of the previous generation of technology platforms, in this instance, especially Microsoft. Even in their infancy, GAI tools are already accelerating and amplifying the radical restructuring of our news gathering and publishing, online communications, and advertising systems set into motion by the rise of Google, Meta, Amazon, Apple, and Microsoft. For example, GAI is already facilitating the creation of “a new generation of misinformation sites” powered by AI bots, capable of attracting paid advertisements and undermining legitimate sources of information.

To understand the task before us, it is important to put GAI into a full historical context. The key lesson is that the technology is largely wielded – and indeed dominated – by a handful of tech companies that have already imperiled the news media by intentionally disrupting the economic models that historically supported public interest journalism. Google and Meta siphon revenues from news outlets through in a variety of ways. They scrap and display content developed by other companies and individuals on their own platforms, restricting or even preventing users from sharing news with one another, and punishing or extorting publications by diverting readers away from news sites. This also includes rolling up control over the infrastructure for digital ads and exploiting this position to divert billions of dollars in advertising away from supporting journalism into their own vaults. Moreover, they have built business models based on the personalized manipulation of the news, information, conversations, and commercial offers presented to individuals, which has allowed a variety of actors to exploit these platforms to transmit disinformation and propaganda designed to manipulate debate, public opinion, and voting, as well as purchasing.

In short, Google and Meta grew into the tech behemoths they are today largely by starving news publishers of funds that would otherwise have been used to support journalism.
Enforcing copyright laws and requiring companies to meet their legal obligations — whether that is compensating rights holders or ensuring transparency of training data — are critical in order to protect human creativity and reward the labor that goes into it.

**The Rollout of GAI Benefits Platform Monopolists at the Expense of the Public Interest**

Generative AI depends on vast amounts of data. This means that underpinning all of GAI is — in a play on the title of Justice Louis Brandeis’s well-known book describing how monopolist’s concentrate power, *Other People’s Money* — other people’s content. Without the scraping and cataloging of the words and images that authors, performers, journalists, and artists have made available online, these GAI tools *would not* be possible. They are merely digitized assemblies of infinitely complex forms of human expression that exploit the labor of hundreds of millions of people without consent, attribution, or fair compensation for private profit.

Consumer-facing GAI requires enormous energy, resources, technological infrastructure, and, most importantly, data. As such, only the most dominant technology corporations, such as Microsoft, Amazon, Apple, Google, and Meta – with trillions of dollars in market capitalization – are able to develop and deploy the foundational models upon which GAI tools are built and the cloud infrastructure on which they depend. For example, the latest publicly available GAI model for ChatGPT reportedly cost more than $100 million to develop and requires almost a million dollars a day to operate.

News and media, specifically, are among the most important sources of information for training models and are critical for real-time AI-powered searches. In an important Google dataset that is used to train some of the most popular large language models (LLMs), including some by Google and Meta, news makes up half of the top 10 sites in the training data. GAI foundation models include content that was put behind paywalls and intended to be restricted to paid users. In September, ChatGPT and Bing had to stop a new product partnership because users were able to bypass publisher paywalls.

Should the current rollout of GAI applications continue in this legally uncertain environment, or were copyright protections granted to these models and/or their outputs, it will create perverse incentives that will further accelerate the dominance of Big Tech over core components of AI infrastructure. Moreover, given these corporations’ ability to rapidly integrate AI into the vast array of products and services they already own, GAI is already becoming yet another way for Microsoft, Amazon, Apple, Google, and Meta to extend and entrench their already dominant market positions. This will make it difficult, if not impossible, for sectors like journalism or the creative industries to remain independent, much less to maintain a public interest orientation in the case of the news industry.

The current implementation of GAI also furthers efforts by the platform monopolists to extend their dominance over our information channels and our public discourse. This in turn will exacerbate the ways in which these corporations are already threatening and cheapening real journalism and exploiting the labor of millions of journalists and politically engaged individuals
to build their models and develop valuable applications. This includes by harming publishers’ revenues and worsening disinformation.\textsuperscript{23}

The ultimate effect will be to amplify and accelerate the many ways in which the concentration of power and control by the platform monopolists are disrupting and undermining democratic debate and democratic institutions in the U.S. and around the world.\textsuperscript{24}

Proposals to “allow” publishers to de-index their sites or pages from crawlers, as with Google’s recently announced robot.txt extender,\textsuperscript{25} is not the answer to protecting their intellectual property much less the broader public interest. De-indexing news from AI models and stripping journalism from training data undermines public interest goals by reducing the supply of quality information and making downstream applications less accurate and reliable. Furthermore, this approach puts the burden on publishers to protect against misappropriation of their content rather than requiring those corporations to obtain permission or license content. The multibillion-dollar companies at the forefront of tech innovation must be required to track their use of rights-protected content, obtain a license, or provide compensation. We do that with most copyright materials and their creators – a radio station or streaming service is allowed to play a song or show a movie if they pay a licensing fee. Artists are not required to pull their materials from platforms to be compensated. This same spirit should animate how we treat copyright and GAI.

Additional considerations:

**Lack of transparency.** Currently, there is no transparency in the source material of these models,\textsuperscript{26} the use of licensing of rights-protected materials, how companies “make decisions” to generate outputs, and the degree and extent of human control and monitoring. These GAI models operate entirely without oversight by either the public or federal or state administrative agencies, and unless they are required to respect copyright (and data privacy) laws, there is little incentive for companies to develop transparency mechanisms for source data. No regulatory regime can adequately advance the public interest without establishing proper mechanisms that enable oversight of how these tools are constructed.\textsuperscript{27}

**Exacerbation of mis/disinformation.** GAI services are being used for illegal or other nefarious purposes. Consider that users continue to be able to circumvent imposed restrictions to generate explicit imagery and content, or to supercharge the ability to disseminate mis/disinformation into our communications channels.\textsuperscript{28} As a result, GAI services are already being used at prolific rates to imitate and impersonate people, fabricate non-consensual pornographic images, and undermine democracy. For example, AI technologies are already being exploited to transform authentic images of lawmakers and interfere with the democratic process.\textsuperscript{29}

**Fast-spreading of consumer harms.** GAI services are also already being used at prolific rates to imitate and impersonate ordinary people – again by creating images and videos that are not just false but also endanger their lives and livelihoods.\textsuperscript{30} In particular, such actions are being weaponized against women as an escalated form of sexual harassment.\textsuperscript{31} GAI services are even being used to replicate people’s voices to scam consumers and trick identity verification systems.\textsuperscript{32}
The rollout of GAI clearly embodies the worst aspects of the “move fast and break things” mantra of the U.S. technology industry. All the aforementioned actions from users, although not always illegal, indicate that powerful technological tools are being rolled out by companies without adequate safeguards against misuse and abuse. Despite this circumstance, GAI are rapidly being integrated into numerous existing consumer products and services, such as our computer operating systems and various smartphone applications. Developers of these technologies, regulators, and Congress must establish guardrails to limit how these tools can be used.33

Recommendations

The Open Markets Institute recommends that any policy recommendations, agency actions, or legislative proposals must incorporate at least the following:

1) Grant copyright protection to the input side of GAI and not copyright protection to GAI output content, which is likely to entrench the monopoly power of Big Tech while continuing to undermine a range of industries that derive their value from creating original and artistic creations. One traditional policy that can help serve as a model was how the U.S. government for most of the 20th century sharply differentiated between individual innovators and large corporations in the enforcement of patent rights.34

2) Any degree of copyright protection granted to GAI outputs must empower and protect human creators, publishers, and artists, by ensuring fair compensation to authors and clear attribution to the source material. Under this approach, we will acknowledge the fundamental power asymmetries between human creation and machine automation under the control of corporate giants.

3) All GAI models should be required to provide “nutrition labels” of the input sources and must be subject to a comprehensive third-party review to ensure compliance with the aforementioned proposals. It will also give the public detailed information relating to the source material referenced for all outputs and the rights to that source material.35

4) Enact policies that facilitate the coordination between authors, artists, trade associations, and unions and developers of GAI technologies to establish fair agreements on the use of content.

Suggested Additional Reading


End Notes

1 The Open Markets Institute is a non-profit organization dedicated to promoting fair markets. It does not accept any funding or donations from for-profit corporations. Its mission is to safeguard our political economy from concentrations of private power that undermine fair competition and threaten liberty, democracy, and prosperity. The Open Markets Institute regularly provides expertise on antitrust law and competition policy to Congress, journalists, and other members of the public. The vigorous enforcement of the antitrust laws against corporate mergers and unfair competitive practices is essential to protecting the U.S. economy and democracy from monopoly and oligopoly.

2 In this comment letter, we address the following questions as detailed in the notice of inquiry: Questions 1, 3, 15. Artificial Intelligence and Copyright, 88 Fed. Reg. 65205 (issued Sept. 21, 2023).


5 U.S. CONST. art. 1, § 9, cl. 8.

6 Nicholas Garcia, Generative AI is Disruptive, But More Copyright Isn’t the Answer, PUB. KNOWLEDGE (May 11, 2023), https://publicknowledge.org/generative-ai-is-disruptive-but-more-copyright-isnt-the-answer/.


14 LOUIS BRANDEIS, OTHER PEOPLE’S MONEY AND HOW THE BANKERS USE IT (1914).


16 Sarah McQuate, *Q&A: UW Researcher Discusses Just How Much Energy Chatgpt Uses*, UW NEWS (July 27, 2023), https://www.washington.edu/news/2023/07/27/how-much-energy-does-chatgpt-use/ (quoting Sajjad Moazeni as stating that ChatGPT likely uses “around 1 GWh each day, which is the equivalent of the daily energy consumption for about 33,000 U.S. households.”); Matt O’Brien & Hannah Fingerhut, *Artificial Intelligence Technology Behind Chatgpt Was Built In Iowa — With a Lot of Water*, AP NEWS (Sept. 9, 2023), https://apnews.com/article/chatgpt-gpt4-iowa-ai-water-consumption-microsoft-f551fde9803d17a7e8d904f8be822c4 (stating every query on ChatGPT uses nearly one 16 oz water bottle for every 5 to 50 queries); Alex Hughes, *Chatgpt: Everything You Need to Know About Openai’s GPT-4 Tool*, SCI. FOCUS (Sept. 25, 2023), https://www.sciencemag.org/future-technology/gpt-3 (stating that GPT-3.5 was trained off of 300 billion words, amounting to over 570 GB of data “obtained from books, web texts, Wikipedia, articles and other pieces of writing on the internet.”).


21 Laurie Sullivan, *Browse With Bing In ChatGPT Integration Disabled Because Of Users Bypassing Paywalls*, MEDIA POST (July 5, 2023), https://www.mediapost.com/publications/article/386924/browse-
For 33

Use AI To Mimic Voices Of Loved Ones In Distress

voices

32

F

Aligned Information Operations and The Distortion of The Public Spherei


Unless a company decides to voluntarily make data sources known or they are leaked.


E.g., Panos Panay, Bringing The Power Of AI To Windows 11 – Unlocking A New Era Of Productivity For Customers And Developers With Windows Copilot And Dev Home, Windows Blog (May 23, 2023),
