

**Before the
UNITED STATES COPYRIGHT OFFICE**

In re: Artificial Intelligence and Copyright – Docket No. 2023-6

Comments of IAC INC. and DOTDASH MEDIA INC.

October 30, 2023

Honorable Shira Perlmutter
Register of Copyrights and Director of the United States Copyright Office
U.S. Copyright Office
101 Independence Avenue, S.E.
Washington, D.C. 20559-6000

Re: Comments – Artificial Intelligence and Copyright, Dkt. No. 2023-6

Dear Director Perlmutter:

We write to express the views of IAC Inc. (“IAC”) and Dotdash Media Inc., d/b/a Dotdash Meredith (“Dotdash Meredith” or “DDM”) in response to the Copyright Office’s Notice of Inquiry titled “Artificial Intelligence and Copyright,” 88 Fed. Reg. 59942 (Aug. 30, 2023), Dkt .No. 2023-6 (the “NOI”). We much appreciate the Copyright Office’s engagement with stakeholders on these issues and are pleased to have the opportunity to submit these comments and responses to selected questions from the NOI.

We note at the outset that we are in general agreement with the views expressed by the News/Media Alliance in its comments in response to the NOI. We write separately to address certain matters pertaining to the deployment and use of generative artificial intelligence (“AI”) systems that are of special relevance and concern to us.

I. Introduction

Background. IAC, a publicly traded media and Internet company, has been building and growing successful companies for a quarter century. IAC has created eleven public companies over the years, among them Expedia, TripAdvisor, Ticketmaster, Home Shopping Network, LendingTree, Match Group, and Vimeo. IAC is currently home to dozens of popular media and online brands and services used by millions of people each day, including category-leading companies such as home services platform Angi, publisher Dotdash Meredith, and Care.com, the world’s largest platform for care.

Dotdash Meredith is the largest digital and print publisher in America. DDM has more than 40 brands, including iconic titles such as People, Better Homes & Gardens, Travel + Leisure, Parents, Verywell, Martha Stewart, Food & Wine, The Spruce, Allrecipes, Brides, Real Simple, Investopedia, InStyle, and Entertainment Weekly. Nearly 200 million people each month trust and depend upon Dotdash Meredith brands – which have built loyal customer bases over the course of a century – to help them make critical life, health, wellness, parenting, personal finance, and spending decisions, as well as to inform, enlighten, and entertain them.

Creating content that can shape people's important personal decisions, and doing so at the scale required to serve the American people, is an extraordinary responsibility. For this reason, DDM invests heavily – in both human talent and infrastructure such as test kitchens, photo and video studios, and product testing labs – to ensure that the content we create accurately informs and inspires our audience while keeping them out of harm's way.

To take but one example, DDM's health website, Verywell, hosts 15 million unique visitors per month. Our medical experts review every article before it is published on the site, ensuring that each medical fact is correct and that the content is thorough and contains the latest evidence-based guidelines. A team of 20 editors scrubs and updates the site's entire corpus of (today) 24,800 expert-written articles on an ongoing basis (with each article updated at least every two years) to ensure that all Verywell medical content continues to reflect the latest clinical research and guidelines. Our Verywell Medical Expert Board – consisting of 165 of the nation's leading physicians in specialty areas such as cardiology, endocrinology, and obstetrics/gynecology – oversees the medical review and the accuracy of every article.

GAI's Threat to the Internet's Information Ecosystem. Generative AI ("GAI") represents an important, and potentially revolutionary, technological advance. Though still in its early days, GAI's great potential is already apparent, and the technology is being embraced by companies in many sectors of the economy, including Dotdash Meredith.

But GAI also presents great risks that must be addressed – and quickly, given the rapid pace of technological development. We are concerned here with one of those risks: the existential threat that GAI poses to digital publishing, and by extension to the health and vitality of the Internet itself as society's predominant information ecosystem, and ultimately to the public that it informs.

The Internet's ecosystem depends on the free flow of user traffic. Producing quality content is expensive, and as noted above publishers like DDM routinely invest significant capital in the creation of original, trustworthy, usable, and enlightening online content. This content is accessed predominantly through search, which provides users ready access to content creators' websites. Website traffic is these creators' lifeblood, as it results directly in revenue, and in particular advertising revenue – the main, and often sole, source of economic recompense for many publishers since much of the quality content on the Internet is free to users, thereby ensuring equality of access to reliable information. The revenue from user traffic fuels the investments required to maintain content quality.

GAI threatens to upend this balanced ecosystem. GAI companies begin by "ingesting," i.e., copying, enormous swaths of the information on the Internet – staggering amounts of original content created by countless individuals – in order to train the large language models that power GAI. This copying is done without the permission of the creators or owners of that content, and the GAI companies pay nothing for it, regardless of whether the works they copy are protected by copyright. This is essentially theft on a massive, unprecedented scale.

But wholesale copying on GAI's "input" side is only the beginning. GAI's "outputs" also threaten to foreclose content creators from realizing the value of their own works, and thereby recouping the investment necessary to create that content, by choking off website traffic. Specifically, GAI is being developed and deployed by the search engines that act as the gatekeepers to the Internet and play a pivotal role in how Americans discover and access online content. Search engines are taking those vast troves of stolen content and repackaging it, again without permission

or compensation. In the GAI version of search, search engines use the original content of others to answer users' queries directly. The long-familiar links to trusted, authoritative websites where users may access that original content are replaced by unsourced, unattributed, synthetic "answers" based on that content. Even if the answer does identify some original sources, the user often has little or no reason to consult them, as attribution to the brands that publishers have spent millions of dollars and many years building perversely serves to validate the GAI output. And the problematic GAI use cases are by no means limited to search; original content that is copied without permission can be deployed without compensation in many other contexts, from GAI-enhanced personal assistants on mobile devices to GAI-imbued word-processing applications.

The Consequences If GAI's Threat Is Left Unaddressed. If GAI companies are permitted to continue training their large language models by copying content without permission or payment, and if they are permitted to use that very content to divert users from its original sources, the economic incentives necessary for the creation and publication of high-quality original content will wither and die. Publishers and others will not be able to pay content creators enough to produce quality content because they will not receive a sufficient return on investment. Poorer content will reduce website traffic, resulting in reduced revenue, and thus less spending on content creation, spawning even poorer content, and so on in a downward spiral for quality content creators and publishers alike. And all of this could happen very quickly. The Copenhagen Institute for Future Studies has estimated that up to 99% of Internet content could be AI-generated by as early as 2025, in which case "the Internet would be completely unrecognizable." See <https://futurism.com/the-byte/ai-internet-generation>.

This unrecognizable Internet will severely erode user trust. Today, when a user sees Verywell content in the search results when seeking the latest medical guidance on breast cancer treatment, she knows she can trust it. But GAI systems do not, and cannot, offer such assurance. To the contrary, GAI's answers to user queries are generated from a proverbial black box – no one really knows where the answer actually came from, and so no one can tell how trustworthy it really is. And when the "answer" falsely attributes important but inaccurate information to a trustworthy source such as Verywell, the reputational damage to the publisher is exceeded only by the potential harm to the user.

GAI also jeopardizes both people's livelihood and the broader public interest. Deploying algorithms to take the proprietary content of others for free and use it to compete in the marketplace against the actual creators and owners of that content will kill content creator jobs. American ingenuity today is primarily expressed and developed on the Internet, which has become the democratizing societal forum in which the best and the brightest become known and succeed. If the meritocracy of quality online content is replaced by whatever GAI happens to cull from wherever on the web, media in all of its forms will shrivel, copyright will protect nothing of value, and the overall quality of publicly available information, and of our public discourse, will suffer.

Some of those rushing to bring GAI to market dismiss these concerns as mere ammunition in a battle between large companies over dollars. But these concerns are real, and ignoring them would be irresponsible. Reckless deployment of GAI threatens the promising utility of the technology itself. GAI has great potential to enhance the lives of people everywhere, but without quality content for inputs, the quality of GAI outputs will suffer. Indeed, GAI already is prone to so-called "hallucinations" (known to most of us as falsehoods) – statements of fact that are made with complete assurance but that are completely wrong. As actual people become increasingly

absent from the content supply chain, such unreliable GAI output will proliferate online, with potentially dangerous consequences.

The “Innovation” Smokescreen. To all of this – the massive copying without permission or payment, the displacement of original creators and publishers through the misappropriation of their content, the threatened destruction of a vibrant information ecosystem of immense societal benefit – the GAI companies intone the mantra of technological “innovation.” IAC is no stranger to technological innovation, having built leading companies in the evolving Internet space for the past 25 years. Dotdash Meredith, the publishing industry leader in building profitable online brands, fosters innovation, and has embraced GAI, using it to hone decisions about content and to enable advertisers to reach audiences more attuned to their products and services. But there are things that GAI cannot do and that only people can. GAI can’t test products, or taste the dish made from a recipe, or travel the world to find the best getaways. GAI has no judgment, nor even true experience. These unique attributes of human creators, without which there can be no innovation, ought not, and need not, be cast aside to make way for GAI.

Nor is “innovation” a free pass to violate the rights of others. Recall Napster, the peer-to-peer music file-sharing service launched in 1999. It was a textbook example of technological innovation, but it also facilitated copyright infringement on a massive scale, was shut down in 2001, and filed for bankruptcy a year later. Napster was not allowed to ignore the copyright laws in the name of innovation – that would have destroyed the artists on which Napster relied. GAI should be viewed in the same way.

IAC and Dotdash Meredith have been able to achieve success without taking things that don’t belong to them. Presumably the well-funded GAI companies could do so as well. But currently, as they race into a market that Bloomberg Intelligence projects to reach \$1.3 trillion within a decade, they aren’t playing by the rules. Instead, they’re taking other people’s intellectual property and then using it to displace them. In copyright terms, there is nothing “transformative” about any of this, and certainly nothing fair about the use they are making of others’ original content. It’s simply against the law, nothing more.

The Founders themselves recognized that without protecting the work of content creators, sustainable innovation is not possible. The Constitution specifically grants Congress the power to enact laws “to promote the progress of . . . useful arts, by securing for limited times to authors . . . the exclusive right to their respective writings.” Congress accordingly has constructed a statutory regime that protects content creators against unauthorized copying of their works. We urge the Copyright Office to ensure that the promise of GAI is not twisted into a justification for the systematic copying and misappropriating of others’ works, to the great detriment of the authors and publishers who are the intended beneficiaries of this protection.

II. Specific Responses

A. Questions 1 and 2

1. As described above, generative AI systems have the ability to produce material that would be copyrightable if it were created by a human author. What are your views on the potential benefits and risks of this technology? How is the use of this technology currently affecting or likely to affect creators, copyright owners, technology developers, researchers, and the public?

2. Does the increasing use or distribution of AI-generated material raise any unique issues for your sector or industry as compared to other copyright stakeholders?

We support and embrace generative AI technology and its far-reaching potential to help us better understand the world, perform tasks with immense efficiency, solve complex problems, and improve information access. But we believe that these benefits are achievable only if GAI technology is developed and deployed with sufficient guardrails to manage its risks and with due regard for the rights of others.

Unfortunately, fueled by an arms race among competing tech platforms, generative AI products have been prematurely rushed out for public use without sufficient guardrails or regulation to protect either the public at large or content creators and publishers from dangers that the owners of these systems have themselves acknowledged. It is already clear that before GAI is developed further, elected officials and regulators will have to confront and address myriad potential issues with this powerful technology, some already identified, some only glimpsed today, some still unknown.

The known dangers include theft of intellectual property on an unprecedented scale that, as described in the Introduction above, is putting the future of the Internet as a reliable and safe medium at risk.

A related threat is the creation and spread of misinformation and disinformation by malicious (or merely careless) actors that undermine our reputation and the credibility of our institutions. Generative AI chatbots provide search results that substitute for publishers' original content, at times verbatim or otherwise as a detailed summary, without apparent need to consult the original sources. Those chatbots often output inaccurate information (so-called "hallucinations"), including inaccurate information that purportedly comes from our properties but was never published by us. These falsehoods put our users at risk and severely damage our brands' reputations, some of which, like Better Homes & Gardens, have been built over the course of a century.

Imagine, for example, a GAI system answering a young user's prompt for a cake recipe with a recipe that calls for bleach and that it falsely represents as sourced from one of our highly regarded cooking publications such as Allrecipes. Or imagine a GAI answer that falsely states that Verywell recommends a supplement to lower blood pressure when that supplement in fact raises blood pressure. Such GAI misinformation could both seriously harm a consumer and severely damage or destroy the reputation of a publication that has for decades provided (at great cost and effort) accurate and reliable content.

As noted in the Introduction, we are particularly concerned that the major tech platforms that operate generative AI systems and services have significant, even dominant, market power in online search and digital advertising, through which our publishing businesses (and virtually every other digital content creation business) earn the lion's share of their revenue. As currently deployed, the use of GAI (itself trained on copied content) in online search will often obviate users' need to visit the original content creators' websites and interact with their content. This loss of traffic will decimate publishers and other creators, rendering them unable to continue to produce the fresh, accurate, quality content that the public has come to expect and rely upon (and that itself is necessary to adequately train GAI's large language models). The public, too, will suffer, as inaccurate, untrustworthy, and potentially harmful content permeates the Internet.

B. Question 5

5. Is new legislation warranted to address copyright or related issues with generative AI? If so, what should it entail? Specific proposals and legislative text are not necessary, but the Office welcomes any proposals or text for review.

Copyright law as written prohibits the wholesale misappropriation of intellectual property on which generative AI systems apparently depend. Copyright law already prohibits the ingestion and reproduction of copyrighted content for purposes of training GAI large language models, as well as the subsequent reproduction of that content as synthetic output, and we believe the courts will so rule.

We understand, however, that GAI companies somehow have taken a different view, asserting as a defense to infringement that the systematic copying of copyrighted content in order to train GAI systems, as well as the presence of that content in GAI outputs, is protected as a “fair use” of that content. In order to avoid expensive and potentially protracted litigation to confirm that this is a serious misreading of the law, it would be useful for Congress to clarify that the ingestion of copyrighted content for generative AI purposes is not *per se* “transformative” or *per se* a fair use, and that the four fair-use factors codified at 17 U.S.C. § 107 apply fully in the GAI context just as in any other context.

On a related note, we understand that the Copyright Office is working with the News/Media Alliance, Digital Content Next, and other organizations representing rightsholders on simplifying the procedures for obtaining copyright registrations for bulk digital works and websites. Such simplified procedures should prove beneficial in helping us register our digital content, and we are generally supportive of these efforts.

C. Question 8

8. Under what circumstances would the unauthorized use of copyrighted works to train AI models constitute fair use? Please discuss any case law you believe relevant to this question.

8.1. In light of the Supreme Court's recent decisions in Google v. Oracle America and Andy Warhol Foundation v. Goldsmith, how should the “purpose and character” of the use of copyrighted works to train an AI model be evaluated? What is the relevant use to be analyzed? Do different stages of training, such as pre-training and fine-tuning, raise different considerations under the first fair use factor?

8.2. How should the analysis apply to entities that collect and distribute copyrighted material for training but may not themselves engage in the training?

8.3. The use of copyrighted materials in a training dataset or to train generative AI models may be done for noncommercial or research purposes. How should the fair use analysis apply if AI models or datasets are later adapted for use of a commercial nature? Does it make a difference if funding for these noncommercial or research uses is provided by for-profit developers of AI systems?

8.4. What quantity of training materials do developers of generative AI models use for training? Does the volume of material used to train an AI model affect the fair use analysis? If so, how?

8.5. *Under the fourth factor of the fair use analysis, how should the effect on the potential market for or value of a copyrighted work used to train an AI model be measured? Should the inquiry be whether the outputs of the AI system incorporating the model compete with a particular copyrighted work, the body of works of the same author, or the market for that general class of works?*

While the fair-use inquiry can sometimes be a difficult one, involving balancing statutory factors pointing in one direction against factors pointing in another, the massive and systematic copying of copyrighted content for an avowedly commercial and substitutive purpose does not present a hard or close case.

In conducting the fair-use inquiry, it is important to assess the overall context in which the copying is occurring, including the reasons why. Why are companies taking enormous amounts of content from the Internet and using it to train large language models? To what use is all of this information ultimately being put? Is it to compete against and ultimately displace the owners of that content, or for some benign purpose? Thus, the act of ingesting large volumes of Internet content to train a GAI system cannot be judged in a vacuum and dubbed “transformative” merely because it purportedly represents a technological advance. Rather, in assessing a real-world situation, the fair-use analysis must take into account the various ways in which generative AI companies use copyrighted content without consent, and how those uses relate to each other.

For example, in the search context, GAI companies use publisher content in ways that go far beyond training and that serve to displace original content creators, such as retrieving and copying our most recent and most relevant content in order to “ground” generative AI outputs. Search, as it has evolved and as we know it today, is focused on helping people find the best content on the Internet, particularly in cases where the user has high intent to find, research, or plan something specific. In areas such as travel, health, personal finance, food, technology, parenting, and the like, users today have easy access to a spectrum of trusted voices and sources to research online and discover the content they need. GAI-powered search misappropriates that content, separating its owners from its value by separating publishers from their audience. Such siphoning of our website traffic and associated revenue does not constitute a fair use under at least the first and fourth fair-use factors: the purpose and character of the use (including whether such use is of a commercial nature), and the effect of the use upon the potential market for or value of the copyrighted work.

The Supreme Court’s recent decision in *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*, 143 S. Ct. 1258 (2023), is instructive on this issue. In that case, the Court analyzed factor 1 of the fair-use analysis and held that Andy Warhol’s use of Lynn Goldsmith’s photograph of Prince to create a silkscreen portrait of the musician was not a fair use. In reaching this conclusion, the Court relied heavily on the substitutive nature of Warhol’s later work, noting: “the first factor [of the fair use analysis] relates to the problem of substitution—copyright’s *bête noire*. The use of an original work to achieve a purpose that is the same as, or highly similar to, that of the original work is more likely to substitute for, or ‘supplan[t],’ the work.” *Id.* at 1274. The Court then observed that: “If an original work and a secondary use share the same or highly similar purposes, and the secondary use is of a commercial nature, the first factor is likely to weigh against fair use, absent some other justification for copying.” *Id.* at 1277.

When a generative AI system, in response to user prompts or queries, outputs copies or derivatives of publishers’ content, the purpose of that system is the same or highly similar to the

publishers' original purpose – to provide information to users. This theme is central to Justice Gorsuch's concurrence, which stated that the fourth fair-use factor "requires courts to ask whether consumers treat a challenged use 'as a market replacement' for a copyrighted work or market complement that does not impair demand for the original." 143 S. Ct. at 1290 (Gorsuch, J., concurring). In this regard, we note that there is a new and growing market for licensing large corpora such as ours for purposes of AI training and grounding. Several GAI companies have announced licensing deals with content owners for the use of their content for GAI training and building new products; and several companies have launched GAI products trained exclusively on owned, licensed, or public-domain content. But it takes only a few companies with significant market power taking and using publisher content without consent or compensation to eviscerate the value of that content and destroy that nascent licensing market.

D. Question 9

9. Should copyright owners have to affirmatively consent (opt in) to the use of their works for training materials, or should they be provided with the means to object (opt out)?

Developers must obtain express, written, advance consent from copyright owners for use of their copyrighted works for GAI training purposes – period. The burden to obtain such affirmative consent is and remains at all times with the developers who want to use the works; the copyright owners bear no burden to police the unauthorized use of their works. Consent should be presumed denied unless it has been expressly and affirmatively given.

That the content at issue may be publicly accessible on the Internet makes no difference, as website operators' terms of service typically make clear. Someone invited into another's home to enjoy his library is not thereby authorized to take or copy his books.

E. Question 10

10. If copyright owners' consent is required to train generative AI models, how can or should licenses be obtained?

If a generative AI company wishes to use copyrighted material to train a GAI model, it should ask the copyright owner for consent and seek to negotiate a license. That is the ordinary and accepted process for obtaining the right to use the intellectual property of others, and GAI uses are no exception. The increasing number of bespoke licensing agreements between content creators and GAI companies demonstrates that this is a viable approach.

Unfortunately, the major GAI companies have eschewed this approach, opting instead to simply take copyrighted content without notice for undisclosed use in training GAI models. That it now may be burdensome to obtain licenses from the many copyright holders whose works were misappropriated in this way is a problem of these companies' own making. These companies plainly can afford to pay for licenses. The market caps of Google (Alphabet) and Microsoft each exceed \$1 trillion. OpenAI has recently been reported to be in negotiations for additional investments that would value the company at more than \$80 billion.

F. Question 15

15. In order to allow copyright owners to determine whether their works have been used, should developers of AI models be required to collect, retain, and disclose records regarding the

materials used to train their models? Should creators of training datasets have a similar obligation?

15.1. What level of specificity should be required?

15.2. To whom should disclosures be made?

15.3. What obligations, if any, should be placed on developers of AI systems that incorporate models from third parties?

15.4. What would be the cost or other impact of such a recordkeeping system for developers of AI models or systems, creators, consumers, or other relevant parties?

Developers who take others' intellectual property without consent or compensation should be required to generate, retain, and disclose on request records documenting the intellectual property they used to train GAI models. If the taking is determined to be wrongful, these records should be useful to copyright owners in holding developers accountable – for example, in establishing that a taking occurred and determining its extent. Such requirements are not unique when dealing with valuable property, as the banking and securities laws will attest.

G. Question 16

16. What obligations, if any, should there be to notify copyright owners that their works have been used to train an AI model?

As a practical matter, only developers are well situated to know what works were used to train their GAI models. As stated above in response to Question 15, developers should be required to generate, preserve, and disclose records documenting their taking of works to train GAI models. And as stated above in response to Question 9, because consent should be obtained before copyrighted works are used to train GAI models, developers should also be obligated to notify copyright owners whose works already have been used for such purposes.

H. Question 22

22. Can AI-generated outputs implicate the exclusive rights of preexisting copyrighted works, such as the right of reproduction or the derivative work right? If so, in what circumstances?

AI-generated outputs implicate the exclusive rights of copyright owners at least when the outputs of generative AI systems are reproductions or derivative works of the preexisting works. This infringement analysis may apply, for example, where a developer misappropriated copyrighted content to train a GAI system to generate infringing output or where a GAI system or product accesses copyrighted content through retrieval augmented search or other methods.

I. Question 24

24. How can copyright owners prove the element of copying (such as by demonstrating access to a copyrighted work) if the developer of the AI model does not maintain or make available records of what training material it used? Are existing civil discovery rules sufficient to address this situation?

In the absence of records of ingestion of materials for training, copyright owners may be able to show illustrative examples of generative AI systems' outputting copies or derivatives of

their works. These types of infringing outputs may include verbatim content, detailed summarization, or other evidence of access to and use of an original work.

As noted in answer to Questions 15 and 16, if a company is taking content, without the consent of or compensation to the content owner, for use in training a GAI model, that company should be required to generate, retain, and disclose records documenting the intellectual property it has taken. Under existing civil discovery rules, an adverse evidentiary inference against the company may well be appropriate where it fails to maintain such records.

J. Question 25

25. If AI-generated material is found to infringe a copyrighted work, who should be directly or secondarily liable—the developer of a generative AI model, the developer of the system incorporating that model, end users of the system, or other parties?

In view of the rapidly evolving nature of GAI technology and the varying factual scenarios in which the infringement question may be presented, the developer of the GAI model, the developer of the GAI system, end users (who may include corporate actors), and potentially other parties may be liable for infringement of copyrighted works.

K. Question 26

26. If a generative AI system is trained on copyrighted works containing copyright management information, how does 17 U.S.C. 1202(b) apply to the treatment of that information in outputs of the system?

It is wrong for a company to take a tag off a product and sell it as its own. The same is true here under Section 1202(b), and companies should face liability when they remove copyright management information as part of GAI training or the provision of GAI output.

* * *

In conclusion, we thank the Copyright Office for the opportunity to submit the foregoing comments.

Respectfully submitted,



Kendall Handler
EVP and Chief Legal Officer
IAC Inc.



Jeffrey Hartwig (Oct 30, 2023 19:05 EDT)

Jeffrey Hartwig
SVP and General Counsel
Dotdash Media Inc.