October 30, 2023

Raptive, Inc
1411 Broadway, 27th Floor
New York, New York 10018

United States Copyright Office
Washington, DC

To Whom it May Concern:

We write to you today to provide our comments regarding the Copyright Office's request for public comment associated with The Office's undertaking of a study of the copyright law and policy issues raised by artificial intelligence ("AI") systems. We understand that the Office seeks comment on these issues, including those involved in the use of copyrighted works to train AI models, the appropriate levels of transparency and disclosure with respect to the use of copyrighted works, and the legal status of AI-generated outputs.

Raptive exists to empower digital content creators to make a sustainable living from their work. We provide them with the ability to monetize their websites through digital advertising, and a host of other services to help build and engage their audiences and build their business capabilities. Our connections to and relationships with these creators run very deep. We are submitting this response because there is a very real risk that unless properly regulated, generative AI platforms will harm small companies that create content (and the companies like Raptive that serve them) and make it impossible for them to earn a living from the online exploitation of their creative works (including via digital advertising and search), to the benefit of larger, better resourced companies that use those small creators' content without their permission.

While there are an estimated 100,000 independent content creators on the web, Raptive's customers are 5,000 of the world's best. These creators started with a passion for food, travel, personal finance, parenting, and many other interests and turned that passion into impressive small businesses, writing and creating some of the internet's best original, human-created content. These are true, bootstrapped American small business success stories. Collectively, these 5,000 creators are a powerhouse– making up the 7th largest digital media network in the United States, reaching nearly 200 million U.S. citizens every month, as measured and reported by Comscore. Individually, these creators are at risk of being ignored and exploited by big technology and AI platforms. Our creators also are diverse, with 58% of all programmatic advertising spend on diverse-owned websites enabled by Raptive.

Raptive is actively advocating on behalf of our creators to protect and empower them through historic changes brought on by AI and to ensure that our creators' businesses will be able to grow and thrive in the next chapter of digital media, which will be heavily impacted by AI. We

recently published an Open Letter stating our position on Generative AI and why it is important to protect content creators from unauthorized use of their content by AI platforms. So far, that letter has received thousands of signatures. In addition, over 1,400 of the websites that we serve have signed a letter allowing Raptive to directly advocate on their behalf to protect their rights against the negative impacts of AI on their content and their businesses. While we directly represent the interests of the creators that we work with, the issues that we will raise here are applicable to any digital creator or publisher that creates and publishes original digital content.

We believe that the use of generative AI in connection with online search services, including the AI generated summarization of creators' digital content, constitutes clear copyright infringement. Such use is particularly concerning when such summaries are provided as part of an internet search service, as it would substitute for and drive traffic from the materials that were wholesale ingested to create the service. That is an **un**fair use, without consent and without compensation. The materially adverse harm that will be caused to these creators cannot be understated, particularly when large search platforms are the single largest source of traffic of people visiting these creators' websites.

As just one example, Google is currently testing a Beta version of its new search generative experience (called SGE) that fully incorporates AI generated content directly sourced from content creators. Members of the public can enroll in the beta at https://labs.google.com/search to experience the beta search experience first-hand. It appears that the result of SGE is that this AI generative content supplants that of the original content creator. In other words, the little blue links that are the source of traffic to these websites will no longer be the main part of the search engine experience, they will become secondary in favor of the AI generated content. This means that the creator's content is being used to create the very content that steals traffic from them.

Google's Beat version of SGE is now using that content to replace that of the creators and significantly reducing the amount of traffic sent to their websites. Based on our own analysis of the likely effect on search traffic to these websites, we estimate that Google's SGE experience will effectively reduce search traffic to creator websites by as much as two-thirds. That reduction will result in catastrophic harm to traffic, revenue and resulting livelihoods of these creators, with many of them shutting down, prompting job losses and the loss of original, human created content for the digital ecosystem.

We believe it is clear that use of creators' content in this fashion violates their exclusive right to prepare derivative works based on their copyrighted content. *See* 17 U.S.C. § 106(2); *see also Warhol v. Goldsmith*, 143 S. Ct. 1258, 1275 (2023). Essentially, these smaller creators are being forced to compete with their own content, to the benefit of free-riding larger companies.

In no way can such unauthorized use be considered a fair use, as every fair use factor would favor Raptive's community of small creators. *See* 17 U.S.C. § 107. Factor one favors them as the use of their content by generative AI platforms is deeply commercial and is substitutive, not transformative. *See*, *e.g.*, *Warhol v. Goldsmith*, 143 S. Ct. at 1276. Factor two also favors

2

Raptive's creator community as their works are creative. *See*, *e.g.*, *TCA Television Corp. v. McCollum*, 839 F.3d 168, 184-85 (2d Cir. 2016). Factor three similarly favors the creators as the entirety of their works is being copied. *See*, *e.g., Harper & Row, Publishers, Inc. v. Nation Enterprises*, 471 U.S. 539, 565-66 (1985); *TCA Television Corp.*, 839 F.3d at 185; *Twin Peaks Productions, Inc. v. Publications Int'l., Ltd.*, 996 F.2d 1366, 1375-76 (2d Cir. 1993). And finally, factor four favors Raptive's creators because, as discussed above, unauthorized use of their content by generative AI platforms harms both the current actual market and potential markets, due to the substitutive use that threatens to devastate these small businesses. *See*, *e.g.*, *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 590 (1994); *Harper & Row*, 471 U.S. at 568; *Fox News Network v. TVEyes, Inc.*, 883 F.3d 169, 180-81 (2d Cir. 2018); *Princeton Univ. Press v. Mich. Doc. Serv.*, 99 F.3d 1381, 1386-87 (6th Cir. 1996); *Twin Peaks*, 996 F.2d at 1377.

We also believe that use of copyrighted content to train AIs likewise violates creators' exclusive rights and is not a fair use. Thus, creators should be compensated for the content that has already been illegally ingested by the AI platforms to train their models, and for any such future use of their content, which should be properly licensed by AI platforms from these creators. Additionally, these creators need to be compensated for the loss of search traffic and associated revenue that will result from the launch of AI-generated search experiences and products.

Put simply, generative AI platforms have the potential to reshape the digital landscape in favor of their owners while cutting out small creators. It is incumbent upon Congress and regulators to understand this story and take appropriate legislative and regulatory action.

Below are Raptive's responses to some of the questions that the Office has posed in its request for comment that are relevant to Raptive's creator community.

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?

While we claim no expertise with respect to use of AI systems in the fields of research, education, biotechnology, and medicine, it may well be that this technology can be a positive force for advancement and innovation in some areas with appropriate safeguards.

However, as we stated in the introduction to this letter, we have serious concerns about the inherent risks that are evident for original, human created content creators, and therefore, copyright owners. The wholesale, unauthorized scraping and ingestion of their digital content, and the obvious use of their content by generative AI in creating derivative works, including summaries, presents clear risks, including the following:

1. The massive ingesting and scraping of copyrighted, human created content by AI platforms threatens the ability for content creators to make a sustainable living from their work. Both the input and the output of large language models (LLM's) is infringing upon creators' work. As mentioned in our introduction, our analysis suggests that the use of generative AI in the online search market will have a massive, negative impact on search engine-driven traffic to

3

creator websites. Replacing links to creator sites with summarizations of creator content produced by generative AI's, supplants the need for consumers to navigate to the very sites that produced the content in the first place.

2. The impact of this market substitution likely will be catastrophic for content creators. Independent content creators rely heavily upon search engines to provide them with audience traffic to their websites. Typically, these websites receive more than 50% of their visitor traffic from search engines, with some receiving as much as 70% of their traffic from this source, so a large-scale decline in traffic would have a materially adverse effects on the ability of these creators to monetize their content with digital advertising, subscriptions, merchandise, affiliate, and other forms of digital monetization.

3. The materially adverse impact to content creators may cause many of them to cease operations. This includes all digital content publishers and creators who create original content.

4. Additionally, the loss of revenue to these creators likely will have a corresponding negative impact on jobs in the space. As we mentioned earlier, many of these creators are small businesses.

5. The proliferation of AI content farms, i.e., LLM synthesized and produced content, at scale has the potential to crowd out access to websites that are built with human created content. The loss of this original creative content will further dilute the quality of the content that we all consume and rely upon as consumers of digital content. Do we really want to trust AI generated content when it is subject to hallucination, particularly at a time when large platforms are already dealing with misinformation and disinformation?

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?

The short answer is "yes," for all the reasons that we have outlined above.

3. Please identify any papers or studies that you believe are relevant to this Notice. These may address, for example, the economic effects of generative AI on the creative industries or how different licensing regimes do or could operate to remunerate copyright owners and/or creators for the use of their works in training AI models. The Office requests that commenters provide a hyperlink to the identified papers.

This piece references the effect on the SEO industry as a result of SGE. This industry is tangential to ours, but it gives a frame of reference to the overall impact that SGE will have on search traffic to creator websites.

This article describes the fight between content creators and AI platforms and search engines over licensing of content for training purposes. We are aware of two examples to date of licensing agreements, one between OpenAI and the Associated Press and another between OpenAI and Shutterstock. In addition, Getty Images announced in September the launch of Generative AI by Getty Images, which "pairs the company's best-in-class creative content with the latest AI technology for a commercially safe generative AI tool."

This study by Search Engine Land offers a methodology for any website to use to calculate traffic loss due to Google's SGE experience, and this piece and this article describe the challenges that digital publishers face from SGE, including lost traffic.

4. Are there any statutory or regulatory approaches that have been adopted or are under consideration in other countries that relate to copyright and AI that should be considered or avoided in the United States? How important a factor is international consistency in this area across borders?

Certainly, the legislative approach that is currently making its way through the approval process in the EU is something that the U.S. should be paying attention to. We support the legislation's proposal that any content produced from generative AI should contain a list of copyrighted training data used to produce that content. The AI platforms argue that they can't effectively provide the copyrighted material used by the LLM's, which in and of itself is a problem for copyright holders. There has been some discussion of opt out provisions for copyright holders, but as we mention in our answer to Question No. 9 below, we would prefer an opt-in provision.

We support the legislation's provisions that require AI platforms to train, design and develop generative AI systems in such a way that there are state-of-the-art safeguards against the generation of content in breach of EU laws, which mainly aim to protect against the infringement of intellectual property rights (and in particular against copyright infringement.)

We support the legislation's provision for substantial fines that apply in cases of non-compliance with the AI Act, and we believe that similar provisions should be considered in any potential US legislation.

The internet operates on a global, cross-border basis, so it is important that businesses and copyright holders are subject to substantially similar and consistent statutory and regulatory approaches when dealing with AI and associated copyright issues. As we have seen with statutory and regulatory approaches to privacy issues in the EU, there is an inconsistent approach to how these regimes are enforced and regulated, which causes uncertainty for businesses when deploying solutions that are meant to comply with the law. Similarly, the US has not adopted federal privacy legislation to address some of the problems that GDPR addresses in the EU.

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.

We strongly believe that the unauthorized use of copyrighted material by generative AI platforms is not subject to the statutory fair use defense. We also believe that the courts ultimately will validate this perspective, but even if they do, it likely will take years for that process to play out. By the time these issues are fully adjudicated, the internet likely will be irrevocably changed, with many content creators potentially losing their livelihoods. To address this danger before the creators are deprived of the ability to generate revenue from their works, we believe that there is a need for legislation mandating explicit consent from creators before their works are used. Creators must have full authority over their works and must be fairly compensated for their contributions. AI systems should not unfairly compete with creators, whose content is fundamental to providing key features, and should duly compensate them. All AI systems must cite their source material, even if it's part of the training data, and dataset transparency is essential–it shouldn't be a secret what data these companies have scooped up. Strict monitoring and enforcement of copyright laws is essential, but currently the burden of enforcement falls on the copyright owners, which is another factor that we think deserves consideration. This legislation should be grounded in our five core principles: acquiring consent, ensuring control, providing fair compensation, citation, and preventing copyright

infringement. Congress must make a clear statement that creative works made by human beings, not machines, are worthy of protection.

We advocate for retrospective compensation for creators whose work has already been used. Furthermore, a robust compensation mechanism should be established to protect creators' interests and livelihoods in the future. This mechanism should be based on the degree of harm caused and should address issues such as loss of search traffic and ad revenue. The primary goal is to ensure that creators are compensated for the use of their creative works and shielded from unauthorized use of their content, particularly where it could be used to unfairly compete against them.

6. What kinds of copyright-protected training materials are used to train AI models, and how are those materials collected and curated?

We will limit our discussion here to content generated by websites that publish human created, copyright protected content. In general, such content includes visual works (including photographs), audiovisual works, and text works (including blog posts, articles, reviews, narrative articles, commentary, opinion pieces, and many more) across a wide range of subjects, including sports, nature, family life, education, food, entertainment and culture, science, technology, consumer products, design, and architecture, and many more.

6.2. To what extent are copyrighted works licensed from copyright owners for use as training materials? To your knowledge, what licensing models are currently being offered and used?

While there have been press reports recently that describe some level of negotiation between digital content owners and AI platforms, details on actual deals that have been struck are scant.

The Associated Press entered into a [licensing deal](#) with OpenAI to give them access to text articles to train their model; however, financial terms of the deal have not been disclosed. It appears, however, that some publishers are [blocking AI crawlers](#). We are aware of one instance in which YouTube has entered into a [licensing agreement](#) with Universal Music for AI generated music, and it is our assumption that it is willing to enter into this agreement given that UMG has shown a willingness to [litigate](#) to protect its musicians, and that the creators represented are high profile copyright owners.

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?

In *Warhol v. Goldsmith*, 143 S. Ct. 1258, 1276 (2023), the Supreme Court stated that the first fair use factor:

> relates to the justification for the use. In a broad sense, a use that has a distinct purpose is justified because it furthers the goal of copyright, namely, to promote the progress of science and the arts, without diminishing the incentive to create. . . . A use that shares the purpose of a copyrighted work, by contrast, is more likely to provide "the public with a substantial substitute for matter protected by the [copyright owner's] interests in the original wor[k] or derivatives of [it]," . . . . which undermines the goal of copyright.

Id. at 1276 (citations omitted).

In conducting this analysis, we believe that a key question is what is the content training the AI to do? If, for example, images are being used to train the AI to generate images, then we believe the purpose of the copying is the same as, or similar to, one of the existing purposes of the creators' content, such that it would be a substitute for, and competitive with, the original content.

Our concern about the use of content to train generative AI's is heightened because we believe it is highly likely that content such as that owned by Raptive's creators already is being used to train generative AI's. This conclusion is supported by this revealing article from the Washington Post.

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?

In our view, any analysis of the fourth factor of the fair-use inquiry, which focuses on "the effect of the use on the potential market or value of the copyrighted work," will necessarily be specific to the facts of any particular case. 17 U.S.C. § 107(4). In analyzing the fourth factor, however, we believe that all three of the possible harms identified in the question may be relevant, depending on the particular factual context presented: (1) harm to the market for particular works, (2) harm to the works of a particular author, and (3) harm to the market for a type or category of works (e.g., news reporting).

Moreover, it seems clear that a market already exists for the use of content to train generative AIs, as shown by the articles linked above in response to Question No. 3. Indeed, by one estimate this market was worth $1.73 billion in 2022 and is likely to grow significantly in the future. Given this, it seems highly unlikely that the unauthorized, uncompensated use of content to train generative AIs would <u>not</u> result in at least one of the three types of harm referenced in Question No. 8, and possibly others than have not yet been contemplated.

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)?

We believe that use of content should not be allowed unless copyright owners consent (opt in), particularly as creators are not given an option to consent or opt out when their content is collected en masse by the AI platforms (unless they resort to blocking software, which we believe often will be an option of last resort, particularly for smaller creators, as it would block creators' content from being included in normal search results as well). Opt in gives creators more control over their content.

9.1. Should consent of the copyright owner be required for all uses of copyrighted works to train AI models or only commercial uses?

All uses, but unquestionably for commercial uses.

9.3. What legal, technical, or practical obstacles are there to establishing or using such a process? Given the volume of works used in training, is it feasible to get consent in advance from copyright owners?

Again, our point here is that the burden is on the AI platforms to figure out a solution to gaining consent, which is what they should have done before taking creators' content in the first place. The

volume and scale of works that they have vacuumed up is of their doing, not that of the content creators.

9.4. If an objection is not honored, what remedies should be available? Are existing remedies for infringement appropriate or should there be a separate cause of action?

One possible solution to this is a regulatory structure that includes fines, or perhaps a compulsory licensing structure similar to the one that already exists for musical works, rather than putting the full burden (and cost) of enforcement on the infringed party by requiring it to file suit in court. Another option is a fast-track court proceeding that can quickly adjudicate and award judgment to the infringed party.

10.1. Is direct voluntary licensing feasible in some or all creative sectors?

Yes, we believe that it is. See the example that we cited above in response to Question No. 3 of the announced licensing arrangement between OpenAI and the Associated Press. In Raptive's case, we believe that it would be feasible for us to negotiate on behalf of our creators to obtain direct licenses with AI platforms, and we have put in place a consent structure with our creators to do that.

10.2. Is a voluntary collective licensing scheme a feasible or desirable approach? Are there existing collective management organizations that are well-suited to provide those licenses, and are there legal or other impediments that would prevent those organizations from performing this role? Should Congress consider statutory or other changes, such as an antitrust exception, to facilitate negotiation of collective licenses?

Because this issue between AI platforms and content creators is nascent and so far, there are limited examples of licenses, we have not yet taken a specific position on this, but are open to the possibility, including an antitrust exemption. One key question is: how will differences in the type and quality of content be represented in these types of license arrangements?

Organizations such as ASCAP and BMI, and the Music Licensing Collective, may be useful models for this type of set up.

13. What would be the economic impacts of a licensing requirement on the development and adoption of generative AI systems?

AI platforms and search engines who use generative AI to replace access to original human created content websites stand to benefit immensely from this technology, using content without consent or compensation and building their business off the backs of creators who stand to lose a significant amount of revenue as a result. The correct question is, what is the economic impact of NOT requiring a license?

Sam Altman has stated that his grand plan for OpenAI is to capture a significant portion of the world's wealth, up to $100 trillion, and that he alone will decide how to distribute it. Original content creators stand to receive no compensation at all, meanwhile, even though their works are a crucial input for what has the potential to be the most lucrative technology ever created.

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?

Yes, to both questions.

15.1. What level of specificity should be required?

For digital content, at minimum: identify the source URL used for the content and the specific content used.

15.2. To whom should disclosures be made?

Disclosure should be made to the content owner and to the consumer of the generative content, as well as made publicly available so that researchers can have access for humanitarian and scientific purposes, but public disclosure will also enable complex analysis of the training data to see how it contributed to generative output.

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

The same obligations as identified in the responses to Questions Nos. 15.1 and 15.2.

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

In all cases they should be notified as stated in the responses to Questions Nos. 15.1 and 15.2.

We thank the Copyright Office for the opportunity to comment on these issues, and it is our conviction and hope that content creators are protected and compensated as we enter a new phase of innovation and advances in artificial intelligence.

Sincerely,

Michael Sanchez

CEO

Raptive, Inc.