**Microsoft**

October 30, 2023

The Honorable Shira L. Perlmutter
Register of Copyrights and Director of the US Copyright Office
United States Library of Congress
101 Independence Avenue, SE
Washington, DC 20540

> **Re:** ***Notice of Inquiry and Request for Comment re Artificial Intelligence and Copyright*** **[Docket No. 2023-06]**

Dear Register Perlmutter:

## I. Introduction

Microsoft and GitHub welcome the opportunity to respond to the U.S. Copyright Office's Notice of Inquiry and Request for Comment re: Artificial Intelligence and Copyright [Docket No. 2023-06]. This inquiry raises important and timely issues, and its impact will extend well beyond the creative and technology industries, to governments, organizations and people who will use AI to achieve more.

Microsoft believes that AI has the potential to improve people's lives in ever-expanding ways. The ability of AI to help advance human knowledge and understanding will lead to improvements in medicine, science, and industry. Organizations and individuals will use AI to innovate, create, obtain critical insights, and address significant societal challenges. AI will power tools that make everyone more productive at work, school, or home. Microsoft is confident in the promise of AI and its capacity to improve the human condition.

We recognize, however, that some artists, writers, musicians, and other creators have questions and concerns around the impact that AI, and especially large-scale "generative" AI, will have on their work and their economic opportunities. These fears echo worries voiced around innovative technologies in the past: the printing press, the camera, the photocopier, the VCR, the internet. Microsoft is working with rightsholders today to understand these concerns and explore ways to help mitigate them.  We have

also heard from other professional artists that generative AI is empowering them to make art more accessible and allowing them to pioneer entirely new artistic mediums.[1]

The important focus of this inquiry is how to achieve a balance, both enabling AI's promise and protecting intellectual property interests. Microsoft's longstanding commitment to fostering innovation and respecting intellectual property rights provides a broad perspective on this balance. This perspective is informed by Microsoft's diverse roles:

**As an author and rightsholder for nearly fifty years:** Microsoft and its employees have created and commissioned millions of copyrighted works representing nearly every subject matter enumerated under Section 102 of the Copyright Act. These include musical compositions and sound recordings, books, articles, blogs, video games, feature films and series, graphic art, photography, paintings, sculptural works, architectural works, and mask works, alongside our well-known software offerings.[2]

**As a technology platform:** Microsoft's digital tools are used by hundreds of millions of users, including individuals, organizations, and enterprises.[3] Our customers use our technology and our platforms to empower themselves and their organizations to achieve more. They create, distribute, communicate, and collaborate. They educate, research, manufacture, and sell. They address major societal challenges and drive breakthroughs in productivity and scientific discovery.

**As a developer of AI technology:** For 30 years, Microsoft has delivered AI breakthroughs in vision, speech, language, decision making and machine learning. Microsoft AI powers billions of intelligent experiences every day through products and services.[4] Microsoft both develops and licenses foundation models.[5] Azure AI is used by over 85% of Fortune 100 companies today. Over 1,000 organizations, including startups, multinational corporations, and educational institutions, are applying the capabilities of Azure OpenAI Service to advanced use cases such as customer support, conversational AI, summarization, and gaining insights from

---

[1] *See* letter from artists using AI, available at https://creativecommons.org/about/policy-advocacy-copyright-reform/open-letter-artists-using-generative-ai-demand-seat-at-table-from-us-congress/.

[2] Microsoft's work in this space includes: i) Tens of thousands of software titles, including some of the world's most popular productivity tools, developer tools, operating systems, and apps ii) Tens of thousands of printed and electronic books (https://www.microsoftpressstore.com) iii) Hundreds of thousands of online technical articles (https://visualstudio.microsoft.com/msdn-platforms/) news stories, online publications iv) Tens of thousands of online videos, v) Short and long form films/series/motion pictures vi) Thousands of video games, representing some of the most popular entertainment titles in the world and vii) Musical compositions and sound recordings.

[3] These include enterprises across every sector, including education, public and private research institutions, local, state, and national governments, and every industry, including creative and entertainment, manufacturing, research, technology, education, public service, healthcare, communications, finance, and many more.

[4] These services include Windows, Xbox, Microsoft 365, Microsoft Teams, Bing, Azure AI, Dynamics 365, and Microsoft Defender. Microsoft subsidiaries, including Nuance, also provide AI services, https://www.nuance.com/en-gb/index.html.

[5] Microsoft models include the Turing model, (https://turing.microsoft.com/) and the Florence model (https://azure.microsoft.com/en-us/blog/announcing-a-renaissance-in-computer-vision-ai-with-microsofts-florence-foundation-model/).

data.[6] Similarly, GitHub's AI tools allow developers across the world to create the next generation of their code more efficiently, increasing productivity that enables developers to solve increasingly more complex problems.

In each of these roles, Microsoft is invested in ensuring that copyright laws remain balanced. Microsoft is concerned that disrupting copyright's carefully crafted balance could stifle innovation. A pivotal reason for United States leadership in technological and creative innovation over the last fifty years is because the U.S. has struck the right balance in its copyright laws, protecting intellectual property rights while promoting creativity and invention. We believe that our copyright laws are flexible enough to maintain this balance as AI technology continues to develop and evolve.

## II.     Copyright Protects Expression, not Ideas or Facts

The constitutional and foundational principles underlying U.S. copyright law guide Microsoft's perspectives: copyright law is intended to "promote the progress of science and useful arts." Copyright law has never protected facts, ideas, concepts, or functional elements in a copyrighted work. 17 U.S. Code § 102(b) ("In no case does copyright protection for an original work of authorship extend to any idea, procedure, process, system, method of operation, concept, principle, or discovery"). This idea-expression dichotomy has been a cornerstone of U.S. copyright law for decades and is fundamental to U.S. innovation, enabling everyone to learn from and build upon the ideas and concepts of others. Copyright law has always permitted humans to read and examine copyrighted training materials to learn how to write, understand scientific patterns, or how to paint or take a photograph. AI models, like humans, similarly "learn" patterns, correlations, facts, and methods from ingesting training materials.

The fair use principles embodied in Section 107 of the Copyright Act have also been a cornerstone of the United States's leadership in technology for decades. United States courts have consistently allowed reproduction of a copyrighted work when it is not used to communicate the copyright owner's original expression to the public, but instead the facts, ideas and concepts within the work are used to create something novel. Altering this careful balance, by adding new rights or extending existing ones, would risk the constitutional objectives that our intellectual property systems are designed to achieve.

## III.      AI Will Help Advance All Human Knowledge (Q1-4)

As the Notice of Inquiry's first question acknowledges, AI's impact is much broader than merely the ability to generate works that would be copyrightable if they had been produced by a human author. The varied applications of AI touch every part of our lives

---

[6] Learn more at https://www.microsoft.com/en-us/ai/azure-customer-stories-ai.

and our economy, driving economic growth,[7] enhancing productivity[8] with the potential to have positive labor market impacts,[9] and supporting the labor market in aging populations.[10] Examples of AI's impact can be seen across Microsoft[11] and our customers'[12] adoption in a diverse range of industries.

Microsoft's AI tools are designed to benefit everyone at every organizational level, in all sectors, across the world. The following examples illustrate how generative AI is incorporated into some of Microsoft's products and services already:

> **Microsoft 365 Chat**[13] is a copilot for work that combines the power of large language models (LLMs) with a customer's universe of data in Microsoft 365 apps and provides them with customized and contextual assistance on a diverse range of tasks. These tasks include routine steps like summarizing a document, identifying action items from a meeting, or creating draft emails.

> **Bing Chat Enterprise**[14] is a copilot for web search, which includes the ability to find information using a conversational interface with both text and images.[15]

> **GitHub Copilot**[16] is an AI-powered pair programmer trained on billions of lines of code that turns natural language prompts into coding suggestions across dozens of languages.

Microsoft and customers across a wide range of sectors are also using AI tools to solve real world problems by analyzing and understanding processes, methods, information, facts, and insights contained in documents, media, data, and articles (some of which include copyrighted works). Examples include:

> **Humanitarian Aid**: AI models trained on vast quantities of images from public and private sources are enabling first aid responders to identify and respond to the most urgent situations.[17]

---

[7] https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html#:~:text=As%20tools%20using%20advances%20in%20natural%20language%20processing,by%201.5%20percentage%20points%20over%20a%2010-year%20period.

[8] https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#business-value.

[9] https://www.brookings.edu/articles/how-ai-powered-software-development-may-affect-labor-markets/.

[10] https://www.sciencedaily.com/releases/2021/09/210916114542.htm.

[11] https://news.microsoft.com/2023/04/04/ai-in-action/.

[12] https://www.microsoft.com/en-us/AI/customer-stories.

[13] https://www.youtube.com/watch?v=ZtXZV41zQZA.

[14] https://www.youtube.com/watch?v=6ZG2cdYr1RY.

[15] https://www.youtube.com/watch?v=rd9mYTcT91A&t=6s.

[16] https://resources.github.com/copilot-demo/.

[17] https://www.forbes.com/sites/irenebenedicto/2023/08/19/an-ai-model-tested-in-the-ukraine-war-is-helping-assess-damage-from-the-hawaii-wildfires/?sh=5e651d363754.

**Climate change**: AI tools use large scale AI models to analyze photographs and video from many different sources to help identify locations most vulnerable to climate related disasters, enabling access to early warning systems.[18]

**Healthcare**: AI is enabling rapid drug discoveries and new cures by assisting with analyzing immense amounts of medical and scientific research and reducing the amount of time to digest this data from decades to days.

**Governments**: The public sector is using AI to improve public services, meet the needs of citizens and improve national defenses and cybersecurity using AI. AI models trained on images and video are used to manage traffic, improve safety in town centers and improve urban life, and AI models trained on online information are connecting people to social, recreational and employment opportunities.

**Education:** AI "copilots" assist educators in the classroom[19] to address variability in student learning[20], giving specialized tutoring, and ensuring equal access to education.

IV.     **AI Models are Trained on Vast Amounts of Data**

Access to data is critical for AI development, and recent groundbreaking advancements in AI require the ability to train on vast amounts of training material. Yet, we recognize that copyright holders have questions about how their content is used and legitimate concerns that the output of an AI system may be substantially similar to their own work. A clear understanding of the LLM training process provides important context for addressing these concerns, and for Microsoft's position that the training of an AI model does not constitute copyright infringement. The distinction between training an AI model versus examining the output of an AI system using the model is crucial.

A. Training Materials and Datasets(Q6)

The breakthroughs in the development of large-scale AI (or foundational) models such as Open AI's Chat GPT require training using vast amounts of data. Large amounts of varied data are essential to allow the AI models to perform accurately and without bias.

There is no single approach to the collection, storage, preprocessing, and use of data for training. Developers take different approaches depending on the model, constraints of the development environment and the intended use of the model.  The need to preprocess and curate the data for training has changed over time and continues to change as approaches to technology develop. For example, for some machine learning methods it has been necessary to label and curate specific types of data, making these

---

[18]  https://www.linkedin.com/pulse/using-ai-advance-early-warnings-all-initiative-lavista-ferres%3FtrackingId=6cHTyFuCiUecghlPOinLLQ%253D%253D/?trackingId=6cHTyFuCiUecghlPOinLLQ%3D%3D.
[19] https://www.youtube.com/watch?v=fdh_1GDDW7U.
[20] https://www.youtube.com/watch?v=fHZdcLxdzFQ.

methods labor and resource intensive. Self-supervised methods of machine learning, which do not require a human to label the data, have vastly increased the scale of data that machine learning methods are able to read, giving rise to the increased performance that we are seeing now. Developers of large-scale AI models therefore optimize for the quantity of data, since the more data available to train a model, the better the performance of the model.

In general, materials for AI training are either collected directly by AI developers from public sources or obtained from third parties or rights-holders. Common sources of AI training data for developers include non-profit organizations such as the Common Crawl and The Pile, who play an essential role in making data from publicly accessible sources accessible for training and using AI. This is particularly important for smaller organizations and researchers that do not have the means to generate or collect massive and varied datasets themselves. AI developers may also negotiate data access and sharing arrangements directly with rightsholders or publishers, providing access to data that is not publicly available, or very specialized data. Microsoft is in current discussions with rightsholders to explore such partnerships.

Despite these different means to collect data, there is still a significant lack of access, which prevents many businesses and organizations from developing or benefiting fully from AI. Open approaches to data can alleviate this problem. Many governments are now taking steps to increase access to data, some through legislation.[21] Industry is recognizing that there is a need for greater access to data to drive societal and economic benefits.[22] There remains a critical need to make data more open, not closed.

B. <u>Model Training to Learn Concepts and Patterns (Q7)</u>

From a simple classifier, used in many everyday applications of AI, to large-scale AI models that will power many future applications, AI models learn by identifying patterns, correlations, and concepts across the training data. This process enables new insights from patterns to be gleaned that could otherwise take a lifetime to uncover.

AI models perform mathematical operations. A large language model is a highly complex algorithm with billions of parameters. Since AI models are algorithmic functions that read numbers, not text, words are transformed into "tokens" that are represented as numerical vectors. These vectors are generated to represent not just words but information about the semantic and contextual meaning of the words and their relationships to other words in the vocabulary. This enables the model to correlate relationships between words.

During training, algorithms are trained so that they get better at performing a particular task. For a LLM, training involves improving the model's ability to predict missing words

---

[21] https://eur-lex.europa.eu/eli/dir/2019/1024; https://ec.europa.eu/commission/presscorner/detail/en/ip_20_2102.
[22] https://www.industrydataforsociety.com/.

from sentences it has never seen, based on concepts that it learns. The model stores what it has learned by updating the parameters of the function, referred to as "weights". To do this, the tokenized training data is read by the model. In an example of self-supervised learning, the data will have some tokens masked, i.e., blanked out. This enables the model to predict the missing tokens, and then remove the mask to determine if it predicted the tokens correctly. At the start of the training process the weights may be set randomly and the initial predictions may work poorly. But the model weights will be updated depending on how accurately the model predicts the blanked-out tokens. As more tokens are seen by the model, the model will continue to learn by updating the weights reflecting patterns and trends which relate to underlying concepts in the training data. It is these patterns and trends that relate to concepts that are stored in the model, not the training data.

The tokenized training data is different from the text that was read, and it is used for a distinct non-expressive purpose. The model does not use the original training materials or the tokenized training data to provide its responses. Instead, once trained the model can make predictions to guess outputs based on the patterns and concepts it has learned.[23] These models do not "copy" or "look up" data from a database. They are not recalling and outputting text from a webpage that was contained in the training materials. Fundamentally, they are tools that analyze data to understand patterns so that they can guess outputs. At their core, AI models are extremely advanced and complex statistical models.

V.      **Fair Use Permits Use of Copyrighted Materials to Train AI Models (Q8)**

Microsoft believes that the fair use doctrine in the U.S. is the Intellectual Property framework best suited to supporting AI development. And while there are policy arguments in favor of using copyrighted works in the training of LLMs, the law is also clear that the use of copyrighted material is fair use. We welcome the steps that many jurisdictions are taking, under their national IP frameworks, to clarify that copyright law continues to permit AI training.[24] Countries that provide the greatest clarity to support AI development will enable the greatest adoption of responsible AI technology. Without clarity, no company will be able to confidently develop AI systems. Smaller companies in particular will be most affected because they cannot afford to litigate to establish their fair use of training materials.

---

[23] Examples demonstrating that machine learning models learn rather than memorize is provided in the paper *Sparks of Artificial General Intelligence: Early experiments with GPT-4* https://arxiv.org/pdf/2303.12712.pdf. The authors seek to demonstrate the model's capacity to apply knowledge and skills across different contexts or disciplines to demonstrate comprehension of ideas. Here the authors ask the model to generate an output that it would not have seen in the training data by deliberately picking combinations of domains that the training data would rarely include, such as literature and mathematics. One such example includes an instruction to "write a proof that there are infinitely many primes, with every line that rhymes" the model handles this task impressively, demonstrating that it has learned the concept of prime numbers.

[24] This includes Japan, Singapore, Korea, the European Union, and Israel.

A. <u>AI Models are Transformative Uses</u>

Fair Use's first factor asks whether the reproduction of a copyrighted work is "transformative" – does it add something new with a different purpose or character"?[25]As the Supreme Court has instructed, the larger the difference in purpose, the fairer is the use.[26]

Making an intermediate copy of a work to discern the underlying concepts, methods, facts, and patterns to train an AI model is completely different from copying an expressive work to communicate the copyright holder's original expression. Just as humans are permitted to learn from the ideas, concepts, and patterns in copyrighted materials, copyright law has for decades recognized that fair use principles allow intermediate copying and use of copyrighted materials for the purpose of learning and creating new, transformative works. Courts have even allowed the reproduction of libraries of copyrighted works in enabling technical processes, because the resulting new work is highly innovative and serves a very different purpose from the original copyrighted works.[27]   As a result, the first fair use factor weighs in favor of allowing copying for the purpose of training an AI model.

B. <u>Training an AI Model Does Not Affect the Value of the Original Work</u>

The potential market for, or value of, a copyrighted work is not affected by use of the work to train an AI model. If a person has legal access to a copyrighted work, the value of the work is not diminished if the person analyzes and learns from the work. This is the case whether the work is read directly, or if it is analyzed using tools in the course of training an AI model. Any suggestion that this impacts the value of the work would extend the copyright owner's exclusive rights in ways that control the non-expressive, unprotected elements.

Authors and artists may justifiably be concerned over the potential impact of outputs of an AI model that are substantially similar to their copyrighted work. While Microsoft and other AI developers take steps to mitigate the risk of such outputs, they can potentially occur. In these circumstances, existing United States copyright law can provide a remedy. The creation of an infringing output, though, is distinct from the training of a model to learn patterns and concepts. For example, a person can know all the lines of a copyrighted poem and receive inspiration from it on structure and meter, but that knowledge does not reduce the value of the work and it does not infringe. Infringement only occurs if the person then publishes a new work that is substantially similar.

---

[25] *See Campbell v. Acuff-Rose Music, Inc.,* 510 U.S. 569, 579 (1994).
[26] *See Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith* (2023) (Slip Opinion at page 12). *Warhol* (Slip Opinion at page 16).
[27] *See* Authors Guild, Inc. v. HathiTrust, 755 F.3d 87 (2d Cir. 2014) (a searchable digital database is transformative fair use); Authors Guild, Inc. v. Google, Inc. 804 F.3d 202 (2d Cir. 2015) (reproductions of books as part of a searchable database used to analyze, understand, and extract knowledge is fair).

C. License and Consent Requirements would Limit Competition and Inhibit Technological Progress (Q9)

Any requirement to obtain consent for accessible works to be used for training would chill AI innovation. It is not feasible to achieve the scale of data necessary to develop responsible AI models even when the identity of a work and its owner is known.[28] Such licensing schemes will also impede innovation from start-ups and entrants who don't have the resources to obtain licenses, leaving AI development to a small set of companies with the resources to run large-scale licensing programs or to developers in countries that have decided that use of copyrighted works to train AI models is not infringement. Moreover, without access to a broad set of training materials from varied sources, AI models may become biased or inaccurate. Other proposals that practically limit access to content, such as an opt-in or license system, suffer the same drawbacks.

While there is no basis in copyright law to prevent the use of publicly available work for AI training, voluntary industry standards do provide publishers with control over the collection of their online content for this purpose. Machine readable tools such as robots.txt have long given online content publishers the ability express an intent to "opt out" of data collection for AI uses. And AI developers generally respect these requests without imposition of legal requirements to do so. Microsoft recently heard feedback from some publishers that they preferred more flexibility in these tools, so that they could opt-out of collection for use in AI model training while remaining in other services like web search. Microsoft developed new web controls responsive to this feedback.[29] It is essential, though, that "opt outs" not discriminate among AI developers, a practice that could inhibit competition and innovation.

VI.     **Generative AI Outputs; Opportunities and Protections for Creators**

Microsoft is willing to work with artists, authors, and other content creators to understand concerns and explore possible solutions. We have adopted and will continue to adopt various tools, policies and filters designed to mitigate the risk of infringing outputs, often in direct response to the feedback of creators. Some important questions go beyond the boundaries of copyright law, asking about the impact AI will have on the future of art and business models of creative work. This impact may be independent of whether copyrighted works were used to train a model, or the outputs are similar to existing works. We are committed to ensuring that artists and authors can learn how to incorporate and use AI in their work so that they share in the benefits of this technology. We are also open to exploring ways to support the creative community to ensure that the arts remain vibrant in the future.

---

[28] https://academiccommons.columbia.edu/doi/10.7916/D8XP7G2N/download.

[29] Microsoft recently announced new options for webmasters to control usage of their content in Bing Chat *see*, https://blogs.bing.com/webmaster/september-2023/Announcing-new-options-for-webmasters-to-control-usage-of-their-content-in-Bing-Chat.

A. <u>Copyright Protection Should Extend to Creators Using AI(Q18)</u>

As demonstrated in the U.S. Copyright Office's listening sessions and in a letter to members of Congress,[30] the creative industries are eager to embrace the opportunities AI creates. It is important that human authors be able to secure copyright protection in their works regardless of what types of tools they use in their creative process, whether more traditional tools such as cameras and filters, or more technically advanced tools such as computer aided design software or generative AI. Typically, an author will use substantial creativity and judgment to instruct the AI tool to produce the desired result. For decades, authors have used both human and technical assistants to create their works, particularly for large works such as architectural designs or massive murals, but use of those assistants and tools have never blocked authors from copyrighting their works. The author still controls the creative process and decides on the finished creation.

Consider for example, software developers that are using generative AI to assist in the generation of code. GitHub Copilot is behind an increasing percentage of lines of code written by developers using the tool, 46% early this year, and predicted to increase to 80% in the coming years.[31] However, the developer is in control of the entire development process: the structure of the program, how they prompt Copilot for suggestions, how they accept, iterate on, or edit suggestions. Developers are able to be more creative when using Copilot because they remain focused on the creative process rather than searching for documentation and examples. For this reason, the current guidance from the Copyright Office to set a threshold level of creativity, and only claim the human contributions, is not feasible to follow.

As a policy matter, disallowing copyright protection for works created with the assistance of AI tools unreasonably limits the beneficial uses of these tools for artistic creation. Individuals who use these tools as part of their creative process will need certainty whether the works they generate will be eligible for copyright protection. Without such assurances, the commercial viability of the works made using AI tools is undermined. The adoption of these tools will also be impacted.

B. <u>AI Developers and Users Must Respect Copyright (Q22-23)</u>

When a person uses an AI application to create expressive works, it is possible that AI generated outputs may infringe copyright if the output is substantially similar to a previous work. Just like using any other general purpose tool, such as a photocopy machine, camera, computer, smart phone, users must take responsibility for using the tools responsibly and as designed. When users employ AI tools to create allegedly

---

[30] https://creativecommons.org/about/policy-advocacy-copyright-reform/open-letter-artists-using-generative-ai-demand-seat-at-table-from-us-congress/.

[31] https://github.blog/2023-02-14-github-copilot-now-has-a-better-ai-model-and-new-capabilities.

infringing works, courts should undertake the same considerations they would in any other infringement analysis—evaluating factors like substantial similarity and fair use.

To address the concerns of rightsholders, AI developers have taken measures to mitigate the risk of AI tools being misused for copyright infringement. Microsoft incorporates many such measures and safeguards to mitigate potential harmful uses across our AI tools. These measures include meta-prompts and classifiers, controls that add additional instructions to a user prompt to limit harmful or infringing outputs. For example, Bing Chat will decline to provide song lyrics or provide extracts from books that are available online. The operation of meta-prompts and classifiers are further explained in Microsoft's white paper, Governing AI: A Blueprint for the Future.[32]

Microsoft has also introduced new options for webmasters to control use of their web content in responses provided from Bing Chat. Using this feature, results that are identified in search can be blocked from being provided through the chat interface. This change came out from collaboration with rightsholder communities. And Microsoft has offered the ability for living artists to request that their name not be used to generate prompts.[33] These steps are not requirements of copyright law, but Microsoft is committed to listening to the concerns of artists and creators and looking for ways to address potential concerns that arise from the use of generative AI.

Microsoft continues to improve current mitigations and implement new ones in response to our learnings and encourages rightsholders to help us think through effective industry best practices.  GitHub's recently announced reference feature was developed with engagement and feedback from the developer community. It lets developers choose whether to block code that matches code in public repositories or allow the code suggestions with information about the matching public code on GitHub, further placing developers in the driver's seat when using these tools.[34]

Microsoft has also committed to indemnifying and defending customers of our commercial Copilot offerings if a third party sues for using Microsoft's commercial copilot offerings or the output generated by these tools, provided that the customer has used the guardrails built into the products.[35] This Copilot Copyright Commitment reflects Microsoft's commitment to building responsible, AI-powered products and tools that limit the risk of infringing outputs. It also provides a strong incentive for Microsoft customers to adopt responsible practices to mitigate these risks. This program helps Microsoft educate users on appropriate uses of AI technology and reinforce how users can respect intellectual property rights.

---

[32]  https://blogs.microsoft.com/on-the-issues/2023/05/25/how-do-we-best-govern-ai/.
[33] https://www.bing.com/images/create/help?FORM=GENHLP.
[34] https://github.blog/2023-08-03-introducing-code-referencing-for-github-copilot/.
[35] https://blogs.microsoft.com/on-the-issues/2023/09/07/copilot-copyright-commitment-ai-legal-concerns/.

C. Protecting Creators and the Public through Labelling (Q28)

We have listened to the concerns of creators and understand that there is a desire from some creators to be able to identify when a work was created with the assistance of an AI tool. We also understand that there is a clear public interest in being able to identify disinformation and deep fakes. Industry efforts to address labelling are ongoing. Microsoft is a steering member of the Coalition for Content Provenance and Authenticity (C2PA) and incorporates this standard into many products. We are also signatories of the Partnership on AI's Responsible Practices for Synthetic Media,[36] and we are eager to work with stakeholders to ensure that industry is supportive of the needs identified that require labelling.

It is important to carefully consider how AI is used and will be increasingly used, both by the creative industries and elsewhere before adopting legal requirements for labeling. For example, AI, like other software graphics tools that have been used for decades, is being used to create components of creative works, such as backgrounds in film and gaming, and elements of songs. AI may also be used in the creative process to help create first drafts and provide inspiration for plots. It is important to ensure that we consider these scenarios when we consider obligations on labelling since these will be increasingly commonplace. The likelihood that something will be purely AI generated without human involvement is highly unlikely. However, as the use of the technology develops, we are open to supporting new requirements as they emerge.

VII.  **Microsoft Supports Efforts to Find a Federal Solution to Name, Image, and Likeness (Q30-Q34)**

Copyright law provides robust protection and essential remedies for creators whose protected expression is infringed by others, including when AI tools are used to create infringing works. Similarly, trademark laws protect consumers and give creators meaningful remedies when third parties attempt to unfairly pass off goods and services as being authorized by a creator. However, given the capabilities of AI to generate convincing replicas, there is a potential need to provide clarity and certainty at the federal level for protection of digital replicas—unauthorized depictions of names, images, likenesses, and signatures, created without the consent of the depicted artist and used in commerce.

Presently, protection against such unauthorized uses is provided by a patchwork of state statutory and common law rights. The scope and duration of those rights can vary widely, and to the extent in conflict with federal copyright law, ambiguity may exist for whether and when state rights are preempted. Similarly, statutes may lack clarity in providing key exemptions and limitations to enable legitimate, constitutionally protected uses exercised across a range of diverse media and online platforms. Just as copyright and

---

[36] https://syntheticmedia.partnershiponai.org/.

trademark law provides a clear foundation protecting legitimate interests of their owners while permitting the public to exercise legitimate permissionless uses, so too would a harmonized federal law provide much needed clarity in protecting against unauthorized and illegitimate commercial uses that depict names, images, likenesses, and signatures of designated individuals, while also enabling legitimate expression.

In developing a federal law, Microsoft supports an approach that convenes a broad group of stakeholders, and examines how successful, balanced approaches, with carefully scoped statutory rights, terms of protection, and critical exceptions, have worked in states including California, Washington, and New York. A federal statute need not necessarily cover or displace well-developed state law precedent, but it should inform what protections and legitimate uses apply to digital depictions created by AI, and how those rights co-exist with copyright and trademark protections as well as permitted uses.

## VIII.    **Conclusion**

We appreciate the time and consideration the Copyright Office has given to this issue. For the reasons stated above, Microsoft believes that any changes to the Copyright Act would inhibit innovation and would disturb the delicate balance of our copyright laws. However, Microsoft takes its responsibility in this space seriously and has and will continue to adopt mitigations to protect the rights of creators and against any harms to this community. In addition, Microsoft has previously supported right of publicity legislation and is open to considering a bill on the federal level to protect against the unauthorized commercial use of a creator's name, image, and likeness.


Respectfully submitted,

/Burton Davis/

Burton Davis
Vice President and Deputy General Counsel, Intellectual Property Group
Microsoft Corporation