



Via email: AIPartnership@uspto.gov

Re: Comments in Response to USPTO Request for Comments on Intellectual Property Protection for Artificial Intelligence Innovation

Dear Director Iancu:

Adobe, Inc. ("Adobe") submits the following comments in response to the USPTO's notice entitled "Request for Comments on Patenting Artificial Intelligence Inventions," published in 84 Fed. Reg. 58141 (Oct. 30, 2019) (the "Request for Comments").

I. BACKGROUND

Adobe has been setting the standard for digital experiences since the company's founding in December, 1982, and is a global leader in digital experience and digital media solutions, offering products and services that allow creative professionals to create, deploy, and optimize digital content. Adobe prides itself on providing its users, who range from emerging artists to global brands, with the tools they need to design and deliver exceptional digital experiences.

Adobe's business is comprised of three cloud-based solutions – Adobe Creative Cloud, Adobe Document Cloud, and Adobe Experience Cloud – and all of them use Artificial Intelligence ("AI") to deliver intelligent features across Adobe products.¹ As a core element of our platform, we use AI, machine learning, and deep learning capabilities to tackle today's complex experience challenges. Adobe has many examples of innovations powered by AI that are used by creative professionals, knowledge workers, and enterprise customers. Adobe's AI offering, which we refer to as "Adobe Sensei," powers intelligent features within each of our cloud offerings to dramatically improve the design and delivery of digital experiences, blending the art of content with the science of data.

In the creative space, Adobe is focused on using AI to deliver useful tools to the creative professionals who use our products every day. For instance, when graphic designers use an Adobe tool to create an image in a creative work, they can use AI-assisted search in Adobe Stock to find stock images on criteria like emotions (search for "happy") or abstract concepts (search for "jump" or "love"),² filmmakers can review footage and have edits suggested to them, and other creative professionals can have Adobe Photoshop customized to their skill level and work focus.³ For this reason, Adobe maintains a unique

¹ <https://theblog.adobe.com/amplifying-human-creativity-with-artificial-intelligence/>.

² <https://theblog.adobe.com/adobe-sensei-ai-ml-image-search-discovery/>.

³ <https://theblog.adobe.com/demystifying-and-democratizing-artificial-intelligence/>.

position because it uses AI technologies to create tools for creative professionals who then use those tools to produce copyrightable works.

As noted in USPTO's Request for Comments, "Artificial Intelligence (AI) technologies are increasingly becoming important across a diverse spectrum of technologies and businesses." As a highly innovative company,⁴ Adobe has embraced AI to enhance its user experiences and empower its customers to create and innovate.

AI opens creative possibilities, but truly realizing these possibilities requires reasonably unrestrained access to data⁵ (often including copyrighted content) to train artificial intelligence systems that enhance products across our business and enhance our users' ability to more efficiently innovate and unleash their creativity.

Our response primarily addresses two of the questions in USPTO's Request for Comments:

Q3: To the extent an AI algorithm or process learns its function(s) by ingesting large volumes of copyrighted material, does the existing statutory language (e.g., the fair use doctrine) and related case law adequately address the legality of making such use? Should authors be recognized for this type of use of their works? If so, how?

Q9: 9. How, if at all, does AI impact the need to protect databases and data sets? Are existing laws adequate to protect such data?

We will respond to both these questions together, as they are closely related, especially in terms of how the answers to these questions can affect access to data, which is critical for society to fully realize the potential of AI. Adobe appreciates the opportunity to provide feedback on these important questions and looks forward to future opportunities to continue the discussion.

II. COMMENTS

A. Artificial Intelligence is critically important in today's world.

Before addressing the specifics of the questions posed, it is important to understand the global significance of AI and to appreciate the critical role that data plays in AI technologies. Vladimir Putin, noted in a meeting with students in 2017, that the development of AI raises "colossal opportunities and threats that are difficult to predict now" and warned that "the one who becomes the leader in this sphere will be the ruler of the world."⁶ In 2017, China published its "Next Generation Artificial Intelligence Development Plan," which laid out plans to ultimately become the world leader in

⁴ Adobe was listed as #13 on Forbes 2018 list of the "World's Most Innovative Companies."

⁵ Note that when we use the term "data" we mean to include copyrighted materials, such as images, that may be included in datasets used for training algorithms to create machine learning models.

⁶ <https://www.cnbc.com/2017/09/04/putin-leader-in-artificial-intelligence-will-rule-world.html>.

artificial intelligence by 2030.⁷ The U.S., which is currently the world leader in AI,⁸ is contemplating how to maintain that role. In February of 2019, President Trump issued an Executive Order on Maintaining American Leadership in Artificial Intelligence which stated that “Artificial Intelligence (AI) promises to drive growth of the United States economy, enhance our economic and national security, and improve our quality of life.” The Order also notes that “Continued American leadership in AI is of paramount importance to maintaining the economic and national security of the United States and to shaping the global evolution of AI in a manner consistent with our Nation’s values, policies, and priorities.”⁹

AI affects all sectors of business and society and is quickly becoming the way we understand and process the vast amounts of information that is available today. However, AI is highly dependent on access to high quality data, and large quantities of it, and if we limit access to data too much, we will cripple the United States’ ability to keep up with other countries innovating in AI and deprive people of its vast benefits.

A number of other countries, who have grasped the importance of AI and the role of data, have adapted their copyright rules to facilitate the development of AI. These are often referred to as text and data mining (or TDM) exceptions. Japan recently amended their Copyright Act to add exemptions for the use of copyrighted works for machine learning. Article 30-4 allows use of a copyrighted work “...where such exploitation is not for enjoying or causing another person to enjoy the ideas or emotions expressed in such work.”¹⁰ And Article 47-4 facilitates machine learning by allowing incidental copies of copyrighted works.¹¹ Other countries, such as Singapore, Australia, China, and Thailand, are also looking to update their copyright laws to further facilitate machine learning. The EU has recently adopted limited TDM exceptions and continues to explore further refinement of these exceptions.¹² As noted above, the United States is a world leader in AI currently, but could quickly fall behind if our policies hinder or limit access to data.

B. Although the fair use doctrine and related case law supports the legality of processing copyrighted material for the purpose of training AI models, specific guidance is critical to provide the certainty that is needed to enable AI to flourish.

There are strong arguments that fair use should allow the use of copyright materials to train AI models. Although there are no cases yet specifically in the area of AI and machine learning, there are cases in areas outside of AI that support these arguments.

⁷ <https://multimedia.scmp.com/news/china/article/2166148/china-2025-artificial-intelligence/index.html>.

⁸ <https://www.techrepublic.com/article/china-aims-to-overtake-us-as-global-leader-in-ai-innovation/>.

⁹ Exec. Order No. 13,859, 84 Fed. Reg. 3967 (February 14, 2019).

¹⁰ Copyright Act (Japan) Article 30-4. <https://wipo.lex.wipo.int/en/text/504411>.

¹¹ *Id.* Article 47-4.

¹² See, e.g., Association of European Research Libraries, Europe Needs a Broad and Mandatory TDM Exception (Nov. 13, 2018) <https://libereurope.eu/blog/2018/11/13/europe-needs-a-broad-mandatory-tdm-exception/>.

For example, in *Sega vs. Accolade*,¹³ the use of a copyrighted video games to gain understanding of unprotected functional elements was held to be fair use. As with training a model for machine learning, no code from the Sega games ended up in Accolade's games, but copies of the games were made in the process. The court found that the commercial use was indirect and not of great significance because the use at issue was solely an intermediate one (i.e., to discover the functional requirements needed for compatibility).¹⁴ The court also considered the public benefit of the use – an increase in the number of independently designed video games available for the Genesis console – and found that this growth in creative expression was precisely what the Copyright Act was intended to promote.¹⁵ In *Sony v. Connectix*,¹⁶ the court found that reverse engineering Sony's BIOS software to develop game emulator software was transformative and constituted fair use. Like with Sega, the court found that Connectix' commercial use was an intermediate one and did not outweigh the extent of transformation. These cases provide strong support that making an intermediate copy of copyrighted material to gain an understanding of unprotected elements for the purposes of training machine learning models constitutes fair use.

There have also been cases in the area of search tools that further support that copying that occurs in connection with creating a trained model is fair use. In *Perfect 10, Inc. v. Amazon.com, Inc.*, the Ninth Circuit found the use of the thumbnails in Google Image Search was "highly transformative," noting: "Although an image may have been created originally to serve an entertainment, aesthetic, or informative function, a search engine transforms the image into a pointer directing a user to a source of information." Similarly, a trained model also serves a different purpose than the original image – to analyze images or other data to draw useful conclusions. In *Authors Guild v. Google, Inc.*,¹⁷ the court held that Google's copying and digitization of entire books for the purpose of providing search, preview, and analytical research tools was fair use. The court noted that "transformative uses tend to favor a fair use finding because a transformative use is one that communicates something new and different from the original or expands its utility, thus serving copyright's overall objective of contributing to public knowledge." Creating trained models to facilitate machine learning certainly can be said to expand the utility of the original images and contribute to public knowledge. Trained models provide a public benefit and promote the progress of science and the useful arts.

These cases provide strong arguments for fair use when copyrighted materials are used merely to train an AI model. However, because there have been no cases yet specifically addressing the issue of fair use in the context of AI and machine learning, and because fair use is a fact driven analysis, companies and research institutions may be hesitant to rely

¹³ *Sega Enterprises Ltd. v. Accolade, Inc.*, 977 F.2d 1510 (1992).

¹⁴ *Id.* at 1523.

¹⁵ *Id.*

¹⁶ *Sony Computer Entertainment, Inc. v. Connectix Corp.*, 203 F.3d 596 (2000).

¹⁷ *Author's Guild v. Google*, 804 F.3d 202 (2d Cir. 2015).

on fair use, at least without specific and clear guidance. There are other potential arguments for allowing the use of copyrighted materials for AI, such as that the limited and temporary intermediate copying and use of copyrighted material for the purpose of training AI is ephemeral or de minimis and, therefore, doesn't constitute an infringement.¹⁸ However, there is also no direct case law applying these concepts in the context of AI and machine learning.

With the uncertainty inherent in applying preexisting legal theories to new technology such as AI, fear of infringement claims often drives companies without clear access to good data to use low quality data, or insufficient amounts of data, to train models for AI. This affects the quality of the models and the ability for all companies to compete equally. As discussed in Section C below, outdated and insufficient data also contributes to algorithmic bias and puts the United States at a disadvantage in the development of AI as compared to other countries that have clarified the law to clearly permit such uses.

C. Allowing the use of copyrighted materials for training AI models serves the goals of copyright, promotes innovation, and significantly benefits society.

Providing guidance establishing a clear right to use copyrighted materials to train machine learning models is consistent with the goals of Copyright Law. The fundamental purpose of copyright law is to “promote the Progress of Science and useful Arts.”¹⁹ This guidance, whether clarifying the applicability of the fair use doctrine or establishing the right to make this use through concepts such as ephemeral or de minimis use, would help encourage further innovation in and around the science of AI. In addition to promoting the progress and innovation of science, AI can also directly promote the arts and creativity by providing tools, such as those made available by Adobe, that assist artists of various types to innovate in their field.

Ensuring sufficient access to large amounts of high-quality data for training AI is also important to combat algorithmic bias. As was noted in a White House whitepaper on AI, “AI needs good data. If the data is incomplete or biased, AI can exacerbate problems of bias.”²⁰ With the growth of AI has come an increase in examples of AI systems reflecting or exacerbating societal bias. This threatens to overshadow AI's technological gains and

¹⁸ The de minimis exception gets its name from the legal maxim “de minimis non curat lex,” or “the law does not govern trifles.” See, e.g., *Gayle v. Home Box Office, Inc.*, 2018 BL 154554, S.D.N.Y., No. 17-5867, 5/1/18 (holding that plaintiff must prove “that the amount that was copied is more than de minimis” and that “de minimis” can either mean “a technical violation of a right so trivial that the law will not impose legal consequences” or it can mean “that copying has occurred to such a trivial extent as to fall below the quantitative threshold of substantial similarity, which is always a required element of actionable copying.”).

¹⁹ U.S. Constitution, Art. I, Sec. 8, Clause 8.

²⁰ 18. EXEC. OFFICE OF THE PRESIDENT, PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE (2016), https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf.

potential benefits. These range from racial bias in facial recognition technologies to natural language processing which embodies sexist stereotypes.²¹

Copyright laws can exacerbate this problem because companies often feel compelled to use low quality, but readily available data, or insufficient amounts of data, to avoid risks of infringement, even though that data and content may embody bias. Amanda Levendowski, a Clinical Teaching Fellow at NYU, wrote an interesting article that explores the ways in which copyright law can negatively influence the quality of AI, and how fair use might be part of the solution.²² Her article notes that “Copyright law causes friction that limits access to training data and restricts who can use certain data. This friction is a significant contributor to biased AI.”²³ The article goes on to say that “If we hope to create less biased commercial AI systems, using copyright-protected works as AI training data will be key.”²⁴ If AI researchers are forced to use older, lower quality data and content for training models, this type of bias will persist. For example, public domain data is readily available, but these are generally works that were published prior to 1923, which, as Ms. Levendowski notes in her article, “was back when the ‘literary cannon’ was wealthier, white and more Western than it is today.”²⁵ In light of these benefits, there should be clear guidance that the temporary processing of copyrighted materials for the purposes of training AI systems is not infringing, either by creating clear exceptions for this type of use, or by clarifying that this is allowed under the fair use doctrine or some other theory.

D. Adobe vigorously supports the ability of creative professionals to protect their work through copyright, and just as humans can rightfully consume copyrighted works to extract factual information and gain insights, so, too, can AI systems for non-expressive use.

Adobe vigorously supports the ability of creative professionals to protect their work through copyright and to realize economic value from their creative works. Adobe has built an industry enabling creative people to express themselves, and has generated thousands of jobs and economic opportunities for creative people to produce excellent work. AI enhances our ability to deliver on the promise of creative tools and our creative professionals appreciate the technological advantages AI can bring to them. The non-expressive use of media to train our AI and, in turn, improve tools for the creative professional is complementary and supportive of their economic model.

²¹ One toolkit that is used to represent how words relate to each other (i.e., Beijing is to China as Warsaw is to Poland) was found to be riddled with gender bias, including a statement that “man is to computer programmer as woman is to homemaker.”

²² Levendowski, Amanda, *How Copyright Law Can Fix Artificial Intelligence’s Implicit Bias Problem* (July 24, 2017). 93 Wash. L. Rev. 579 (2018).

²³ *Id.* at 589.

²⁴ *Id.* at 621.

²⁵ *Id.* at 615.

Humans can read and consume copyrighted materials to extract information and gain insights and learnings from them, and this ingestion by the human brain does not constitute infringement. Likewise, the fact that a temporary copy is made of copyrighted material so that a computer can do the same thing as humans do, should not change this. Where someone is not using the expressive content of a work, but is using it solely to extract factual information, and there is no harm to the economic value of the work, there is no need to compensate authors for their creative expression (which is what copyright protects).

Access to copyrighted materials for training algorithms does not harm the economic value of the materials. Where no copyrighted materials end up in the end product, and the question is solely about the intermediate copying made to facilitate the training of the AI system, the use should either not be considered to be infringing in the first place (i.e., as ephemeral or de minimis), or it should be considered transformative and constitute fair use. If the end product of the use of copyrighted materials does contain copyrighted materials or is substantially similar to the original copyrighted materials, then it might be considered to be an infringing derivative work and a copying of expressive elements of the original work. But this should not affect the question as to whether copies made in the training process itself are infringing. The mere use of copyrighted materials for training is not for the purpose of using the expressive nature of the materials, but rather to gather factual information – to use the material as data in order to understand the parameters and patterns in the data. The law should be clarified to make it clear that this type of non-expressive use does not implicate copyright law or violate the rights of the copyright owner.

III. CONCLUSION

We hope our comments have illustrated the value and importance of AI technology, its furtherance of the goals of copyright law, the benefits to society of securing access to sufficient quantities of data for machine learning, and the limited, non-expressive and non-harmful nature of the use of copyrighted materials for this purpose. We urge that the law be clarified to provide guidance and a clear path for AI researchers to use data, databases, and data sets (including copyrighted materials) for the limited purpose of training AI models, without the fear of liability for copyright infringement. With sufficient guidance, combined with current law, we believe there is adequate protection for databases and data sets. We believe this is essential for the United States to maintain its position as the world leader in the field of AI, and for the public to fully enjoy the benefits from this transformative technology. We would be happy to further discuss this issue if there are any questions or the USPTO requires any additional information.