March 14, 2025

Via email (ostp-ai-rfi@nitrd.gov)

Faisal D'Souza, NCO
Office of Science and Technology Policy
Executive Office of the President
2415 Eisenhower Avenue
Alexandria, VA 22314

Re: Request for Information on the Development of an Artificial Intelligence Action Plan

---

AI Progress appreciates the opportunity to provide comments in response to the Request for Information, issued on behalf of the Office of Science and Technology Policy (OSTP) by the NITRD NCO, which seeks input on the Development of an Artificial Intelligence Action Plan.

AI Progress is a 501(c)(4) organization dedicated to the responsible development of artificial intelligence (AI) tools. The coalition emphasizes the essential role of existing United States copyright law in fostering innovation, protecting intellectual property rights, and securing United States competitiveness in AI innovation. Members include Amazon, Anthropic, Cohere, Google, Meta, Midjourney, Microsoft, and OpenAI.

We share the Trump administration's objective of advancing American leadership in AI by developing policies that not only promote America's AI dominance but also protect against unnecessarily burdensome requirements that hamper private-sector AI innovation. To meet these goals, any AI Action Plan should further AI developers' innovation, foster economic growth, and generate employment opportunities.

The types of computational analysis that underlie AI and machine learning play a central role in delivering on the goals in America's AI Action Plan, from life sciences, energy, security, and financial services to the creative industries. Protecting the right and ability to perform a computational analysis on publicly available data is necessary to achieve those goals. United States copyright law must therefore remain sufficiently flexible to permit developers to gather and analyze varied data sets to make AI models reliable, robust, and effective. Put differently, access to training material is crucial for creating effective AI models, which in turn, supports the development of AI technologies that can solve real-world problems, drive scientific discovery, and create new opportunities for growth and innovation across industries.

Accordingly, it is essential that policymakers not only preserve the flexibility inherent in United States copyright law, but also affirm that it is a protected fair use to train on publicly available works. It is also essential to resist efforts that would impose undue regulatory burdens on AI developers, including legislative efforts on both the federal and state level, which would undermine the advancement of AI. This is all the more critical given the fierce global competition for leadership in the AI field.

Data to Train, Tune & Improve AI Models and Systems is Critical for AI Innovation

AI is a foundational and transformative technology with the power to drive scientific discovery, strengthen national security, and generate new economic opportunities. Effectively leveraging AI is critical for the United States to maintain its position as a global technological leader.

Already, AI is powering economic expansion and industry transformation. Goldman Sachs projects that AI will have a measurable effect on United States GDP by 2027 and may contribute to a 7% annual increase in global GDP over the next decade. Similarly, Bank of America predicts that AI will drive employment growth in fields such as aerospace, information technology, education, and healthcare. As AI technologies continue to evolve, their economic impact will become increasingly pronounced, with the emergence of new industries, businesses, and professional opportunities.

Innovation in AI fundamentally depends on the ability of AI models to be trained on large quantities of data. The technical process of "learning" for AI models means being able to derive patterns, structures, and relationships from across a broad body of content so that they can operate on probabilistic modeling. The amount of data used in training today's AI models improves accuracy and robustness. Attempts to limit or impede the use of data to train AI models risk hindering the United States' technological progress in AI. Policymakers should thus oppose any measures that could lead to a reduction in access to or use of publicly available data for AI training purposes.

Consequently, any attempts to limit or impede the use of data to train AI models risks hindering the United States' technological progress in AI. Other countries have adapted their laws and environments to ensure AI training and models can flourish. Therefore, we make the following recommendations:

- Policymakers should oppose any measures that could block or restrict access to or use of data for AI training purposes; and
- Policymakers should promote measures that permit access to and use of data for AI training purposes.

United States Copyright Law Fosters AI Innovation Through Fair Use

The United States currently leads the world in AI development, thanks, in part, to existing United States copyright law and its robust fair use doctrine. The doctrine of fair use permits unauthorized copying and other uses of copyrighted works where — as with training AI systems — the secondary use is for a new (i.e., "transformative") purpose and does not function as a market substitute for the original copyrighted works.

As was Congress's goal, fair use has played a crucial role in promoting American innovation and technological developments. When deciding to incorporate the doctrine of fair use into the 1976 Copyright Act, Congress recognized fair use as "one of the most important and well-established limitations on the exclusive right of copyright owners," which should be "adapt[ed]" to account for "rapid technological change." H.R. Rep. No. 94–1476 at 65–66 (1976). Congress noted that fair use had "been raised as a defense in innumerable copyright actions over the years," and that there had been "ample case law recognizing the existence of the doctrine and applying it." *Id.* at 65. Additionally, as courts have long recognized, fair use safeguards the essential freedom of expression inherent in the First Amendment, which creates space for various uses of copyrighted works that do not require permission or payment.

Time and time again, courts have applied fair use to prevent copyright from exceeding "its lawful bounds" and obstructing the development and distribution of new and innovative technologies. *See, e.g.*, *Google LLC v. Oracle Am., Inc.*, 593 U.S. 1, 31 (2021) (finding the use of declarations copied from a computer program's application programming interface (API) to be fair where it "further[ed] the development of computer programs" by allowing programmers to use their acquired skills to develop new applications for a new platform); *Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 217 (2d Cir. 2015) (finding fair use where the digitizing of millions of books to enable a search function was transformative and not meant as a substitute for the authors' books); *Authors Guild v. HathiTrust*, 755 F.3d 87, 97 (2d Cir. 2014) (similar; "the result of a word search is different in purpose, character, expression, meaning, and message from the page (and the book) from which it is drawn"); *Sega Enters. Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1522 (9th Cir. 1992) (finding fair use where computer code was copied for the purpose of reverse engineering and studying how to develop new video games that were compatible with an existing game console).

The use of copyrighted materials to train AI models similarly qualifies as a transformative — and thus fair — use. The purpose of AI training is not to extract and reproduce original expression from training data. Rather, the purpose is for AI models to derive patterns, structures, and relationships from across a broad body of content so that they can generate appropriate responses to user prompts based on statistical relationships. It is these patterns — not the training data's protectable expression — on which the models

rely. The training data is extensively transformed and processed using mathematical formulas, and then broken down into numerical representations and embedded in a complex network of parameters. AI developers' use of training data is wholly distinct from the uses of the authors of such underlying data. Models use the data not to copy their content or challenge authors' ability to sell copies of their works, but rather to develop an entirely new and innovative service that, in turn, produces valuable new content — thereby vastly expanding the capacity for human creative productivity.

As discussed, the development of large-scale AI models necessitates training on extensive volumes of data. Access to a broad and diverse dataset is essential for ensuring the accuracy and effectiveness of AI systems. Fair use supports innovation by ensuring that AI developers are able to access and use the necessary training data that AI developers need to continue to build new AI models.

Given the profound implications for the future of AI, we urge policymakers to recognize that the United States doctrine of fair use is an essential component of any legal framework that will govern AI, and that the use of copyrighted materials to train AI models must ultimately be considered fair use as a matter of law. Our Founding Fathers sought to "promote the Progress of Science and useful Arts." U.S. Const. art. I, § 8, cl. 8. Transformative secondary uses like model training further this Constitutional goal by encouraging the development and distribution of groundbreaking technologies.

Secure America's Global Technological Leadership

Despite the flexibility of existing United States copyright law and the fair use doctrine, both U.S. and international regulators are increasingly considering changes that would handicap the development of AI technologies. The measures being discussed would not enhance model accuracy or safety and are unnecessary for the protection of rightsholders, as the current legal framework already addresses legitimate cases of infringement. Hindering AI development due to unnecessary regulation and restrictions would mean relinquishing leadership in AI to other nations, thus undermining the United States' prosperity and security. Instead, policymakers should focus on advancing the competitiveness of United States AI and creative industries by protecting the core provisions of copyright law that offer technology-neutral protections to legitimate rightsholders and innovators.

Restricting access to AI training data would create self-imposed barriers to innovation, positioning other nations to lead in innovation and encouraging developers to capitalize on more permissive regulatory environments. Indeed, some countries have adapted their laws and environments to ensure AI training and models can flourish. For example,

Singapore and Japan both took significant steps to create copyright laws favorable to AI development. Similarly, China is also considering more developer-friendly copyright laws, including not imposing restrictions on training, to provide a competitive edge in AI development.

The introduction of new statutory licensing frameworks in the United States could fundamentally alter the existing copyright system, creating significant obstacles to technological advancement. Any statutory licensing framework relating to AI training is not only inappropriate given fair use considerations, but it would also create unprecedented administrative burdens, harm small businesses and innovators, slow the advancement of scientific progress, and upend a nascent tool being used by hundreds of millions of Americans. While existing statutory licensing schemes may cover millions of works, a similar regime for works used for AI training purposes would need to administer royalty payments for *billions* of works, including fractional, anonymous, and unclaimed works. It is unclear how such a framework is even possible.

In addition, legislation forcing AI developers to disclose their training data would impair American companies' ability to compete within the global marketplace and present genuine security and competition concerns. Revealing details about the inputs associated with AI models may create opportunities for bad actors to exploit the models and their training programs. The selection of inputs, how those inputs are used during the training cycle, and the data preparation process represent the application of technical skill and know-how that model developers treat as confidential information. Forcing AI companies to disclose the contents of their datasets would gift their confidential business information to foreign AI competitors that may not be subject to similar disclosure requirements. Such forced disclosure would effectively handicap United States technology companies and benefit foreign competitors.

Because overregulation risks disrupting United States innovation, policymakers should prevent the enactment of burdensome regulations on American AI developers. Therefore, AI Progress recommends the following:

- Reject the option of a new statutory licensing framework, related to AI training, as unnecessary, unworkable, and anti-innovation; and
- Prohibit efforts by governments to compel companies to transfer, disclose, or provide access to AI-related intellectual property, such as source code, algorithms, and datasets.

Conclusion

As the United States continues to lead in AI, it is crucial that policymakers recognize and embrace the strength and adaptability of existing United States copyright law.

United States copyright law, particularly the fair use doctrine, already provides the necessary flexibility to address many scenarios likely to arise with generative AI. And courts have decades of experience in applying the fair use doctrine to new technologies. Introducing new statutory requirements or restrictions would create unnecessary barriers to progress, hinder economic growth, and weaken America's competitive edge in the global AI landscape.

We are grateful for the opportunity to provide feedback on this RFI and commend the administration's dedication to maintaining American leadership in AI.

To that end, it is vital to uphold the existing copyright framework, which has long supported technological advancement and to resist efforts that would impose undue regulatory burdens on AI developers in this vital stage of innovation.

We look forward to continuing to engage with the Trump administration, NSF, OSTP, and other agencies to ensure that United States policy remains conducive to innovation, economic prosperity, and global AI leadership.

This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in developing the AI Action Plan and associated documents without attribution.