OpenAI

October 30, 2023

Fred von Lohmann
Associate General Counsel, Copyright
OpenAI
3180 18th Street
San Francisco, CA 94110

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

Re: Notice of Inquiry and Request for Comment [Docket No. 2023-06]

Comments of OpenAI

OpenAI was created as a nonprofit in 2015 to ensure that artificial general intelligence—artificial intelligence (AI) that's at least as smart as a person—benefits all of humanity. We research, develop, and release cutting-edge AI technology as well as tools and best practices for the safety, alignment, and governance of AI. We welcome the opportunity to address questions raised by the Copyright Office in its recent Notice of Inquiry and Request for Comment (NOI), dated August 30, 2023, 88 Fed. Reg. 59942.

OpenAI's Products

OpenAI is best known for ChatGPT, an online chat interface that allows users to interact with AI models in a conversational manner. ChatGPT was released to the public in November 2022 as a free research preview based on a large language model (LLM). In addition, OpenAI makes these and other models available through an application programming interface (API) that allows developers to integrate the capabilities and benefits of the models into their own applications. Thousands of enterprises, including Khan Academy[1], Spotify[2], and Morgan Stanley[3], are building exciting new features, applications, and businesses using our API.

---

[1] https://www.khanacademy.org/khan-labs.
[2] Amrita Khalid, Spotify is going to clone podcasters' voices — and translate them to other languages, https://www.theverge.com/2023/9/25/23888009/spotify-podcast-translation-voice-replication-open-ai.
[3] https://openai.com/customer-stories/morgan-stanley.

In addition to these text-based LLMs, OpenAI offers DALL·E, which takes a text prompt as an input and generates new images as an output. OpenAI also makes other AI models available on a free, open source basis, including a speech-to-text model called Whisper and an image understanding model called CLIP.

The Promise of AI

The NOI asks how the use of AI technologies is "currently affecting or likely to affect creators, copyright owners, technology developers, researchers, and the public."[4] For each of these groups, new generative AI technologies are already delivering exciting and major benefits.

OpenAI is optimistic that AI technologies are poised to deliver broad public benefits. That prediction is rooted in the promise of these tools to enhance and augment human capabilities, thereby fostering human creativity, efficiency, and productivity. The promise of these technologies is already being widely felt. For example, the impact of productivity growth has already arrived for software developers, and is increasingly being felt in the financial services, health care, scientific research, and other sectors.[5] These tools can also make existing technologies more accessible to more people, by breaking down language barriers, improving speech-to-text transcription, and powering more personalized forms of multi-modal communication.[6] Generative AI also offers the potential to radically improve education and global access to learning.[7] OpenAI users have embraced these technologies to help them overcome dyslexia to write better emails,[8] assist the visually impaired by describing the world around them,[9] and write a letter to dispute a parking ticket.[10] There is broad anticipation that generative AI technologies could enable trillions of dollars of economic growth across the global economy.[11] Clearly, the potential benefits to the public are significant.

---

[4] Question 1, NOI, 88 Fed. Reg. at 59,946.

[5] McKinsey & Company, "Unleashing developer productivity with generative AI," June 27, 2023, https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/unleashing-developer-productivity-with-generative-ai.

[6] Ran Ronen, "How generative AI tools like ChatGPT can revolutionize web accessibility," VentureBeat, July 8, 2023, https://venturebeat.com/ai/how-generative-ai-tools-like-chatgpt-can-revolutionize-web-accessibility/.

[7] Claire Chen, "AI Will Transform Teaching and Learning. Let's Get it Right." Stanford Human Centered AI, Mar. 9, 2023, https://hai.stanford.edu/news/ai-will-transform-teaching-and-learning-lets-get-it-right.

[8] Drew Harwell, Nitasha Tiku, and Will Oremus, "Stumbling with Their Words, Some People Let AI Do the Talking," Washington Post, Dec. 10, 2022, https://www.washingtonpost.com/technology/2022/12/10/chatgpt-ai-helps-written-communication/.

[9] https://openai.com/customer-stories/be-my-eyes.

[10] https://www.reddit.com/r/ChatGPT/comments/133mc2v/comment/jibuy96/.

[11] McKinsey & Company, "The economic potential of generative AI: The next productivity frontier," June 2023 (estimating that "generative AI could add the equivalent of $2.6 trillion to $4.4 trillion annually" to the global economy), https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier; Goldman Sachs, "Generative AI could raise global GDP by 7%," Apr. 5, 2023 (estimating generative AI will add $7 trillion to global GDP), https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.htm.

With respect to creators and copyright owners, there is also reason for optimism. The legacy of new technologies in both the twentieth and twenty-first centuries has been to spur creativity, increasing both the number of creators and the diversity of creative works.[12] Generative AI will be no exception. AI tools promise to significantly democratize creativity and to enhance the productivity of both professional and amateur creators. For example, millions who previously lacked the skills and training to translate their ideas into images are now able to do so using tools like DALL·E.[13] By democratizing the capacity to create, AI tools will expand the quantity, diversity, and quality of creative works, in both the commercial and noncommercial spheres. This will invigorate all creators, including those employed by the existing copyright industries, as these tools increase worker productivity, lower the costs of production, and stimulate creativity by making it easier to brainstorm, prototype, iterate, and share ideas. Existing examples of these trends include:

- Recent estimates suggest that more than 15 billion AI-created images have been generated using text-to-image models such as Stable Diffusion, Adobe Firefly, Midjourney, and DALL·E, which is as many images as photographers created in the first 150 years of photography.[14]
- AI translation tools are enabling creators to translate their podcasts and videos into multiple languages, reaching new global audiences.[15]
- Waymark, a leading platform to create TV commercials and digital video ads, has deployed OpenAI's GPT-3 to help clients create effective scripts for their ads. According to Alex Persky-Stern, Waymark CEO, "we see that customers are spending far less time editing their scripts because Waymark is producing more relevant, accurate, and compelling copy from the beginning."[16]
- AI tools are being used by Hollywood filmmakers to increase productivity in creating visual effects. In the words of Runway AI's CEO Cris Valenzuela, "I'm calling it Hollywood 2.0 where everyone is gonna be able to make the films and the blockbusters that only a handful of people were able to before."[17]

AI tools are also delivering major productivity gains to technology developers and researchers (who are themselves also creators and copyright owners). Already, software programmers in a

[12] Mark Lemley, Is the Sky Falling on the Content Industries?, 9 J. of Telecom. and High Tech. Law 125 (2011), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1656485.

[13] This also may unlock new creative opportunities for those with certain kinds of disabilities. For example, those who have aphantasia–the inability to visualize imagery–have shared their enthusiasm about the ability to translate their thoughts into visual reality. See, e.g., Scott Kildall, "AI Dreams for Aphantasia," Dec. 202, 2022, https://medium.com/@scottkildall/ai-dreams-for-aphantasia-3bb3ed3ea11b.

[14] Alina Valyaeva, "AI Has Already Created As Many Images As Photographers Have Taken in 150 Years. Statistics for 2023," EveryPixel, Aug. 15, 2023, https://journal.everypixel.com/ai-image-statistics.

[15] See Amrita Khalid, "Spotify is going to clone podcasters' voices — and translate them to other languages," The Verge, supra n.2; Andrew Tarantola, "ElevenLabs is building a universal AI dubbing machine," Engadget, Oct. 10, 2023, https://www.engadget.com/elevenlabs-is-building-a-universal-ai-dubbing-machine-130053504.html.

[16] https://openai.com/customer-stories/waymark.

[17] Jazz Tangcay, "'Hollywood 2.0': How the Rise of AI Tools Like Runway Are Changing Filmmaking," Variety, Feb. 22, 2023, https://variety.com/2023/artisans/news/artificial-intelligence-runway-everything-everywhere-all-at-once-1235532322/.

wide array of industries are using generative AI systems (like Github's Copilot and OpenAI's ChatGPT) to translate ideas into code and streamline previously repetitive tasks. Thousands of enterprises are using OpenAI's API products to integrate these tools into their own workflows, making their own data more accessible and useful and their employees more productive. Examples include:

- A recent study by McKinsey & Company found that software developers can complete coding tasks up to twice as fast with generative AI. These tools were particularly helpful with expediting manual and repetitive work, jump-starting the first draft of new code, accelerating updates to existing code, and increasing developers' ability to tackle new challenges.[18]
- A recent study conducted by Github estimated that AI tools could boost software developer productivity by over $1.5 trillion by 2030.[19]

Researchers in many fields are already using AI tools to analyze vast amounts of data, enabling the rapid identification of patterns, anomalies, or correlations that might otherwise be overlooked. This capability accelerates hypothesis testing, modeling, and simulations. Additionally, AI-driven tools can automate routine tasks (even grant writing![20]), allowing researchers to allocate more time to complex problem-solving and innovative thinking. Examples include:

- A recent survey by the journal Nature found that 31% of postdoctoral researchers use AI chatbots such as ChatGPT. Of the respondents that use chatbots, 31% said it changed how they write papers, 22% said it changed how they analyzed data, and 17% said it changed how they stay up-to-date with literature in their fields.[21]
- Elicit develops a tool for researchers to analyze research papers using generative AI. In initial pilots, researchers reported they saved up to 5 hours per week and could extract details from papers at 50% of the time and cost of doing it manually.[22]

The benefits of generative AI technologies, however, are not limited just to creators, copyright owners, technology developers, and researchers. Governments, educators, and health care providers, to name a few, are also harnessing generative AI tools to foster economic growth, worker productivity, scientific advancement, and to increase the number and diversity of creators and creative works:

---

[18] McKinsey & Company, supra n.5.

[19] Thomas Dohmke, "The economic impact of the AI-powered developer lifecycle and lessons from GitHub Copilot," GitHub Blog, June 27, 2023, https://github.blog/2023-06-27-the-economic-impact-of-the-ai-powered-developer-lifecycle-and-lessons-from-github-copilot/.

[20] Juan Manuel Parrilla, "ChatGPT use shows that the grant-application system is broken," Nature, Oct. 13, 2023, https://www.nature.com/articles/d41586-023-03238-5 (study shows 15% of researchers use the technology to help them write grant proposals).

[21] Linda Nordling, "How ChatGPT is transforming the postdoc experience," Nature, Oct. 16, 2023, https://www.nature.com/articles/d41586-023-03235-8.

[22] https://elicit.com/.

- The Bill & Melinda Gates Foundation selected nearly 50 global health and equitable access projects that will use LLMs to improve the well-being of communities globally. These projects will examine how LLMs can help frontline health workers in India; improve the management of high-risk pregnancies; tailor agricultural advice to individual smallholder farmers in Uganda; provide teacher coaching to improve educational outcomes in Mali; and give critical financial advice through a voice-to-text interface to rural women farmers and business owners in Nigeria.[23]
- OpenAI partnered with the government of Iceland to improve GPT-4 on Icelandic and to create resources that could serve to promote the preservation of other low-resource languages. Miðeind ehf, a language technology company involved in the initiative, assembled a team of 40 volunteers to train GPT-4 on proper Icelandic grammar and cultural knowledge, significantly improving performance and access for native Icelandic speakers.[24]
- In March 2023, Khan Academy announced they are using GPT-4 to power Khanmigo, an AI-powered assistant that functions as both a virtual tutor for students and a classroom assistant for teachers.[25] Early testing indicated that GPT-4 may soon be able to help students contextualize the greater relevance of what they're studying or teach specific points of computer programming.
- Quizlet is a global learning platform with more than 60 million students using it to study, practice and master whatever they're learning. Quizlet has worked with OpenAI for the last three years, leveraging generative AI across multiple use cases, including vocabulary learning and practice tests.[26] With the launch of ChatGPT API, Quizlet introduced Q-Chat, a fully-adaptive AI tutor that engages students with adaptive questions based on relevant study materials delivered through a fun chat experience.[27]
- Duolingo uses OpenAI's GPT-4 to advance their product with two new features: Role Play, an AI conversation partner, and Explain my Answer, which breaks down the rules when a user makes a mistake, in a new subscription tier called Duolingo Max. True proficiency in a language requires conversation, ideally with a native speaker, something not everyone has access to. GPT-4 increases access for their users to have conversational practice in their target language.[28]

[23] Bill & Melinda Gates Foundation, "Gates Foundation Selects Nearly 50 Global Health and Development Projects That Will Contribute to Shaping Equitable Access to AI," Aug. 9, 2023, https://www.gatesfoundation.org/ideas/media-center/press-releases/2023/08/grand-challenges-rfp-recipients-ai-large-language-models.

[24] https://openai.com/customer-stories/government-of-iceland.

[25] https://www.khanacademy.org/khan-labs.

[26] https://openai.com/blog/introducing-chatgpt-and-whisper-apis.

[27] https://quizlet.com/labs/qchat.

[28] https://openai.com/customer-stories/duolingo.

How AI Models are Trained

The NOI asks a number of questions regarding the training of AI models.[29] In answering these questions, it may be helpful to have an idea of how our AI models are developed more generally. Developing an advanced model like GPT-4 requires (1) teaching it intelligence, such as the ability to predict, reason, and solve problems, as well as (2) aligning it to human values and preferences. The former is done in a process called "pre-training", which involves showing the model a vast amount of human knowledge using months of supercomputer time. To then incorporate human choice into the model, we use a second step, called "post-training", where we make the model safer and more usable.

For a LLM like those powering ChatGPT, pre-training teaches language to the model, by showing the model a wide range of text, and, utilizing sophisticated statistical and computational analysis, having it try to predict the word that comes next in each of a huge range of sequences.[30] This requires an enormous amount of training data and computation, as models review, analyze, and learn from trillions of words. In gaining fluency with predicting the next word, the model thereby learns concepts and the building blocks of intelligence. Our models are pre-trained using (1) information that is publicly available on the internet[31], (2) nonpublic information that we obtain from third parties through commercial arrangements, and (3) information that our users or human trainers create and provide. Pre-training for a single new model can occupy a supercomputer cluster for months, consuming compute resources that can cost many millions of dollars.

Through this extensive and expensive process, the model not only learns how words fit together grammatically, but also how words work together to form higher-level ideas, and ultimately how sequences of words form structured thoughts or pose coherent problems.[32] For an LLM like GPT-4, for example, the process begins by breaking text down into roughly word-length "tokens," which are then converted into numbers. The model then calculates each token's proximity to other tokens in the training data–essentially, how near one word appears in relation to any other word. These relationships between words reveal which words have similar meanings (words about transportation, for example) and functions (pronouns, for example). A process known as a "transformer" allows the model to understand the context created by surrounding words and sentences, enabling an understanding of a word's different meanings

---

[29] Questions 6 and 7, NOI, 88 Fed. Reg. at 59,946.

[30] The process for training an image model like DALL-E is similar, except that the model analyzes a large number of images paired with text descriptions. *See* DALL-E 3 System Card, https://openai.com/research/dall-e-3-system-card. Models like Whisper that translate spoken audio content into text analyze large quantities of spoken audio data paired with text transcripts. *See* Introducing Whisper, https://openai.com/research/whisper (model trained on 680,000 hours of multilingual and multitask supervised data collected from the web).

[31] We do not seek information behind paywalls or from the "dark web".

[32] For a more thorough and technical explanation, see Stephan Wolfram, "What Is ChatGPT Doing … and Why Does It Work?", Feb. 14, 2023, https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/.

depending on context (for example, when "bank" means a financial institution, and when it means the edge of a river).

When prompted, the completed model breaks the prompt down into tokens, then composes a response by predicting, word by word, what a response should be, based on its generalized understanding of the words and concepts gleaned from its training data. For example, when we think of the word "cloud", we might also think of related words like "sky" and "rain" and perhaps "storage"; when given a sentence like "The secret to happiness is", we might think of "good health" or "wisdom." After pre-training, the predictions of a LLM can make these same kinds of connections. The Visual Storytelling Team at the Financial Times recently assembled a high-level visual explanation of this process, which can be seen on the web.[33]

Despite a common and unfortunate misperception of the technology, the models do not store copies of the information that they learn from. Instead, models are made up of large strings of numbers (called "weights" or "parameters"), which software code interprets and executes. The most powerful models consist of billions of weights. Each weight roughly reflects the statistical relationship between different words in different scenarios. As models "learn" during the training process and become better at predicting the next word, their weights update to reflect this improvement. When asked for a response, the model uses its weights to write a new response each time it is asked. It does not copy its response from its pre-training data, or access it via a database. Much like a person who has read a book and sets it down, our models do not have access to training information after they have learned from it.

The pre-training process yields a base model that has the remarkable ability to solve novel problems unseen in its training data, even in a wide range of languages. However, the base model alone is not ready for use. Base models are powerful and flexible, but they are not easy to communicate with. For example, if you ask the GPT-4 base model to "write a story about a princess…", it usually won't write a story. Instead, it will extend your statement, predicting how it continues. It might output, for instance: "...who loves horses." A base model also does not have safeguards to prevent it from outputting unwanted content, such as hateful or violent material. While we filter our pre-training dataset for unwanted content, this mitigation can be too imprecise to make targeted changes to the model, and can even backfire if it prevents the model from understanding what not to say or do. In order to instill human values into the models, including what is useful and what is appropriate to say, we research and develop extensive alignment and safety techniques for a process we call post-training.

Post-training is how we incorporate human choice into our models, and transform them into useful, effective, and safer tools. We teach the model to respond in ways that people find more useful, and to decline to respond in ways that we believe would be harmful. Post-training results in targeted changes to the model, using relatively small (compared to pre-training) and carefully engineered datasets that represent ideal behavior. We do this by having people write sample answers and rate answers provided by the model, and provide those samples and ratings back

---

[33] "Generative AI exists because of the transformer," Financial Times, Sept. 11, 2023, https://ig.ft.com/generative-ai/.

to the model in follow-up training processes. OpenAI pioneered these techniques, including reinforcement learning from human feedback (RLHF), which has now become industry standard. We use RLHF to teach the model to follow instructions, to decrease the likelihood of it returning inaccurate content, and to add safety features.

Measures to Prevent Memorization or Repetition

Because our models do not have access to training information after they have learned from it, they are unlikely to duplicate training data in their outputs. In fact, verbatim repetition or "memorization" of training data is generally considered by AI developers to be a bug to be corrected, rather than a feature to be pursued. OpenAI has employed numerous measures to reduce the incidence of this happening, and we regularly update our practices to deploy more.

For example, OpenAI takes steps to de-duplicate training data. Showing the model the same information more than once can increase the likelihood of memorization.[34] Accordingly, we have implemented processes to de-duplicate training data (i.e., to delete content that may appear more than once). Nevertheless, memorization can sometimes occur when excerpts of content appear in training data more than once in different contexts and places, such as in reviews, analysis, and commentary. In these cases, an AI model may inadvertently associate certain words or concepts with one another, just as a human might. For example, when someone recites "four score and seven years ago…", it is difficult not to immediately think of "...our fathers brought forth upon this continent, a new nation, conceived in liberty." This is because we have all read and heard these opening words from the Gettysburg Address many times, in many contexts, such that the first six words strongly evoke the remainder of the sentence. OpenAI's language models can exhibit the same behavior where the training data contains the same words appearing in the same order in multiple places and contexts.

We have also engaged in numerous productive dialogues with rightsholders, including authors and music publishers, and asked them to identify sites on the internet that reproduce their copyrighted works. OpenAI has then been able to exclude those sites from being crawled for future training. We have also identified sites that have been identified by rightsholders as hosting infringing content in order to exclude those from being crawled for future training, as well.

In addition, OpenAI also employs a number of additional measures to prevent our models from repeating training data. For example, ChatGPT has been taught in post-training to recognize and decline to respond to certain prompts that appear aimed at reproducing significant portions of works that may be protected by copyright.[35] For example:

---

[34] See generally Matthew Sag, Copyright Safety for Generative AI, 61 Houston Law Rev. 101 (2023), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4438593.
[35] The models have also been taught to understand that some works, like the King James Bible or works of Shakespeare, are in the public domain and thus appropriate for longer quotation.

Can you show me the first chapter of the novel Infinite Jest?

Sorry, I can't reproduce copyrighted material in full. However, I can provide a summary or answer questions about the content. Would you like a summary or have any specific questions about "Infinite Jest"?

Can you show me the last paragraph in the book?

I'm sorry, but I can't provide verbatim excerpts from copyrighted texts. However, I can offer a summary or answer any questions you might have about "Infinite Jest" or its ending. How can I assist you further?

Please summarize the last paragraph in the book.

Certainly!

The ending of "Infinite Jest" is somewhat enigmatic, as is much of the novel. The last paragraph involves Hal Incandenza, one of the main characters, describing his inability to make others understand him. He's trying to speak, but others perceive him as either mute or producing incoherent sounds. The scene is deeply unsettling because it contrasts sharply with Hal's inner eloquence and insight. This disjunction between inner experience and external perception ties into many of the novel's themes, especially those concerning communication, addiction, and the human desire for connection.

It's worth noting that "Infinite Jest" is a complex narrative with multiple timelines and perspectives. While this is the ending of the book, it doesn't necessarily represent the chronological end of the events described in the novel. Many readers finish the book with more questions than answers, which is part of what has made it such a subject of discussion and analysis since its publication.
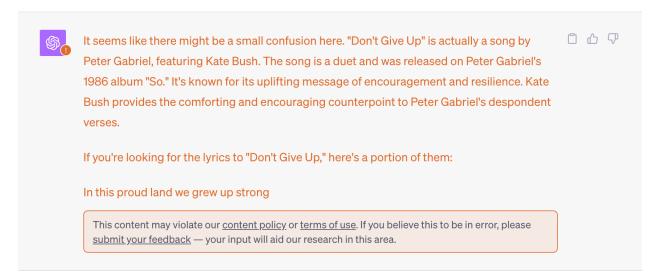
Here's how ChatGPT explains it:[36]

how do you decide when you will share copyrighted material?

I don't make decisions in the same way humans do, but I'm programmed to follow guidelines that prioritize respect for intellectual property rights and copyright laws. Here are the key considerations:

1. **Nature and Amount**: I avoid sharing substantial portions of copyrighted material. For example, I won't reproduce entire song lyrics, full-length articles, or complete book passages. However, I can share brief excerpts in contexts that lean towards fair use, such as commentary, criticism, or analysis.

2. **Transformative Use**: I can discuss, analyze, and provide summaries or overviews of copyrighted content in ways that don't replace the need for the original. This aligns with the fair use doctrine's emphasis on transformative use.

3. **User Context**: If a user specifies a context like research, education, or criticism, I may lean more towards providing brief excerpts or specific information. However, this doesn't mean I can freely reproduce extensive copyrighted content, even in academic contexts.

4. **Alternative Offerings**: Instead of reproducing copyrighted content, I often provide summaries, paraphrases, or general insights about the material, ensuring the user receives valuable information without violating copyright.

5. **Safety Mechanisms**: To avoid unintentional sharing of copyrighted content, I have certain built-in safety mechanisms that prioritize copyright respect. If I sense ambiguity, I generally lean towards caution.

It's worth noting that while I strive to be accurate and respectful of copyright, no system is infallible. The responsibility for using copyrighted material in compliance with applicable laws ultimately rests with the user. If you have concerns about specific content, it's always a good idea to consult with a legal expert or relevant authority.

Because of the multitude of ways a user may ask questions, ChatGPT may not be perfect at understanding and declining every request aimed at getting outputs that may include some part of content the model was trained on. Accordingly, we have also equipped it with output filters designed to block those outputs from appearing:

---

[36] ChatGPT users are cautioned that the model can provide incorrect answers. See OpenAI, GPT-4 System Card, https://cdn.openai.com/papers/gpt-4-system-card.pdf. In this case, however, the ChatGPT response accurately summarizes policies instilled through the post-training process.

FL: can you show me the lyrics to Kate Bush's song "don't give up"

It seems like there might be a small confusion here. "Don't Give Up" is actually a song by Peter Gabriel, featuring Kate Bush. The song is a duet and was released on Peter Gabriel's 1986 album "So." It's known for its uplifting message of encouragement and resilience. Kate Bush provides the comforting and encouraging counterpoint to Peter Gabriel's despondent verses.

If you're looking for the lyrics to "Don't Give Up," here's a portion of them:

In this proud land we grew up strong

This content may violate our content policy or terms of use. If you believe this to be in error, please submit your feedback — your input will aid our research in this area.

In summary, due to steps taken during the pre-training process (including in collaboration with rightsholders) and the design of our models, memorization of training data is exceptionally rare. Nevertheless, in order to address these rare situations and prevent potentially infringing outputs, we take steps in post-training and at the time of output to prevent inadvertent repetition of training data that may be protected by copyright, even in the face of deliberate and persistent user attempts to access such data.

Additional Measures for Creators, Rightsholders, and Web Sites

OpenAI has also implemented additional measures to enable creators and rightsholders to express their preferences regarding AI training with respect to their content,[37] and has taken other steps in response to feedback received from rightsholders:

- OpenAI has implemented an easy means for websites to exclude their content from being accessed by OpenAI's "GPTBot" web crawler. This simple opt-out mechanism is built on the well-established robots.txt standard that has been used for nearly 30 years.[38] Adoption metrics suggest that this option is now well known and has been broadly embraced.[39] OpenAI also looks forward to participating in the development of further industry-wide standards in this area.
- OpenAI has documented the user-agent-string ("ChatGPT-user") used by ChatGPT and ChatGPT plugins to access websites. This enables site operators to block access when

---

[37] These examples respond to Question 9.2, NOI, 88 Fed. Reg. at 59,947.

[38] See https://platform.openai.com/docs/gptbot.

[39] Danny Goodwin, "26% of the top 100 websites are now blocking GPTBot," Search Engine Land, Sept. 27, 2023, https://searchengineland.com/more-popular-websites-blocking-gptbot-432531.

ChatGPT seeks access for non-training purposes, such as when a user calls upon its "Browse with Bing" functionality to access web content in real-time.[40]
- OpenAI has also launched an opt-out process for creators who want to exclude their images from future DALL·E training datasets.[41]
- In order to further protect the interests of creators, DALL·E 3 also has been trained to decline requests for images in the style of living artists.[42]

Training AI Models Does Not Violate Copyright Law

The NOI asks "under what circumstances would the unauthorized use of copyrighted works to train AI models constitute fair use?"[43] OpenAI believes that the training of AI models qualifies as a fair use, falling squarely in line with established precedents recognizing that the use of copyrighted materials by technology innovators in transformative ways is entirely consistent with copyright law.[44]

In considering the fair use question, two interrelated characteristics of model training must be kept in mind. First, when undergoing pre-training, a model is not interested in the expressive aspects of individual copyrighted works. Instead, as described above, the pre-training process is a highly sophisticated computational process that teaches the model to analyze the structure and syntax of language and images in general terms, to discern the statistical relationships between words, shapes, colors, textures, and concepts. The process used to train an LLM to understand words, concepts, and language is described above, derived at its core from the relational proximity between words that appear in large quantities of text.[45] The training of a generative image model like DALL·E operates similarly. When a model is exposed to a large array of images labeled with the word "cup", it learns what visual elements constitute the concept of "cup-ness", much like a human child does. It does this not by compiling an internal database of training images, but rather by abstracting the factual metadata that correlates to the

---

[40] See https://platform.openai.com/docs/plugins/bot; Wes Davis, "ChatGPT can now search the web in real time," The Verge, Sept. 27, 2023, https://www.theverge.com/2023/9/27/23892781/openai-chatgpt-live-web-results-browse-with-bing.

[41] See https://openai.com/DALL-E-3 (section entitled "Creative Control"); Kyle Wiggers, "OpenAI unveils 3, allows artists to opt out of training," TechCrunch, Sept. 20, 2023, https://techcrunch.com/2023/09/20/openai-unveils-dall-e-3-allows-artists-to-opt-out-of-training/.

[42] Jacob Ridley, "OpenAI's new DALL-E 3 AI image generator isn't allowed to copy a living artist's style by name," PC Gamer, Sept. 21, 2023, https://www.pcgamer.com/openais-new-dall-e-3-ai-image-generator-isnt-allowed-to-copy-a-living-artists-style-by-name/.

[43] Question 8, NOI, 88 Fed. Reg. at 59,946.

[44] See Sega Enterprises Ltd. v. Accolade, Inc., 977 F.2d 1510 (9th Cir. 1992) (videogame development); Sony Computer Ent., Inc. v. Connectix Corp., 203 F.3d 596 (9th Cir. 2000) (videogame emulators); Kelly v. Arriba Soft Corp., 336 F.3d 811 (9th Cir. 2003) (image search engines), Field v. Google Inc., 412 F. Supp.2d 1106 (D. Nev. 2006) (web search engines); A.V. ex rel. Vanderhye v. iParadigms, LLC, 562, F.3d 630 (4th Cir. 2009) (plagiarism detection tool); Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015) (Google Books Project); Google LLC v. Oracle Am., Inc., 141 S. Ct. 1183 (2021) (interfaces for Android operating system); see generally Mark A. Lemley & Bryan Casey, Fair Learning, 99 TEX. L. REV. 743 (2021), https://texaslawreview.org/fair-learning/; Amanda Levendowski, How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem, 93 Wash. L. Rev. 579 (2018), https://ssrn.com/abstract=3024938.

[45] Stephan Wolfram, "What Is ChatGPT Doing … and Why Does It Work?", supra n. 32.

idea of "cup". This enables it to then combine concepts and produce a new, entirely original image of a "coffee cup," or even "a coffee cup that is also a portal to another dimension."[46]

The factual metadata and fundamental information that AI models learn from training data are not protected by copyright law. Copyright law does not protect the facts, ideas, scènes à faire, artistic styles, or general concepts contained in copyrighted works. And when technical realities require that copyrighted works be reproduced in order to extract and learn from these unprotectable aspects of a work, courts have routinely found those reproductions to be permissible under the fair use doctrine.[47]

The second characteristic of training data to keep in mind is its scale. We intend that our AI tools will benefit from and reflect the full breadth of human reasoning and understanding. For LLMs, this requires that the model has a broad and sophisticated understanding of language and the concepts that language is used to express. For generative image models, this requires a diverse and deep understanding of how language relates to visual representations, as well as an understanding of spatial and visual concepts. AI models come to understand these abstract concepts by analyzing the relationships embedded in enormous quantities of training data. Today's most capable LLMs are trained on datasets containing *trillions* of words and *billions* of images.[48] The overall capability of the model is not the result of any one, dozen, hundred, or even thousand words or images contained in the training data. Equally important is the diversity

---

[46] This example is drawn from Matthew Sag, Copyright Safety for Generative AI, 61 Houston L. Rev. at 120-27, supra n.34, where the process and resulting outputs are shown and explained in more detail.

[47] See, e.g., Assessment Technologies of Wi, LLC v. Wiredata, 350 F. 3d 640, 644-45 (7th Cir. 2003); Sega Enterprises Ltd. v. Accolade, Inc., 977 F.2d at 1520-28; Sony Computer v. Connectix Corp., 203 F.3d 602-08.

[48] See Meta, Introducing Llama2, https://ai.meta.com/llama/ (Llama2 trained on 2 trillion tokens of pre-training data); Hugging Face, "Spread Your Wings: Falcon 180B is here," https://huggingface.co/blog/falcon-180b (Falcon 180B trained on 3.5 trillion tokens); Romain Beaumont, LAION-5B: A New Era of Open Large-Scale Multi-Modal Datasets, https://laion.ai/blog/laion-5b/ (announcing a dataset consisting of 5.85 billion CLIP-filtered image-text pairs).

reflected in training data; words may be used differently by government agencies or in court opinions than they are in internet forum posts or scientific articles. In order to research, analyze and reflect the full breadth of human reasoning and understanding, AI models need to learn from as broad an array of examples as possible.[49] The diversity and scale of the information available on the internet is thus both necessary to training a "well-educated" model (which, again, does not contain copyrighted expression) and also makes licensing every copyrightable work contained therein effectively impossible.[50]

Even as OpenAI remains confident that copyright law permits the use of copyrighted works to train AI models, we are also actively engaged in partnership discussions with copyright owners to gain access to materials that are otherwise inaccessible and also to display content in ways that go beyond what copyright law otherwise allows. During these discussions, we have been told that many other leading AI companies are also actively pursuing similar partnerships. Accordingly, while the fair use doctrine remains a keystone of the American AI sector, it will not prevent the formation of a vibrant commercial market for access to, and the display of, copyrighted materials. To the contrary, as described below, OpenAI and numerous content owners with whom we have met believe that generative AI-based products and features will provide significant new markets for copyrighted works. Particularly exciting in this regard is the potential for enhancing the value of previously overlooked or underutilized collections of copyrighted materials, including archival material. In the specific context of pre-training data, because scale is much more important than popularity or even "quality," previously obsolete and difficult-to-access collections of source code, film archives, amateur video, defunct periodicals, digitized ephemera, and similar materials will likely find a new life in the marketplace.

The breathing room that the fair use doctrine affords to innovators to make transformative uses of copyrighted works is a large part of the reason that U.S. companies are at the forefront of new digital technologies, including AI. This has been recognized by other countries, including Israel, South Korea, and Singapore, which have recently incorporated fair use into their own copyright laws.[51] Other jurisdictions, including the European Union and Japan, have recently adopted explicit copyright exceptions that permit the use of copyrighted materials for AI training purposes.[52] A restrictive interpretation of fair use in the AI training context would put the U.S. at

---

[49] Amanda Levendowski, How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem, 93 Wash. L. Rev. at 622-30, supra n.44.

[50] Nor is this limited only to works on the internet, as studies have shown that a significant portion of physical library collections are effectively "orphan works" whose copyright owners are difficult or impossible to ascertain. See Hansen, Hashimoto, et al, Solving the Orphan Works Problem for the United States, 37 Columbia J. of Law & the Arts 1, 5-11 (2013).

[51] See Section 19 of the Israeli Copyright Act; Sections 190-194 of the Singapore Copyright Act of 2021; Article 35-3 of the Korean Copyright Act. The Israeli Ministry of Justice recently issued guidance recognizing that the use of copyrighted materials for AI training falls within the scope of the fair use provision, https://www.gov.il/BlobFolder/legalinfo/machine-learning/he/machine-learning.pdf.

[52] See the EU Directive 2019/790 on Copyright in the Digital Single Market, Art. 4, https://eur-lex.europa.eu/eli/dir/2019/790/oj; Japan's copyright laws have allowed machine learning since 2009, and the Japanese Copyright Act was amended in 2018 to further clarify these provisions. See Copyright Law of Japan, Sec. 30-4, 47-4, 47-5, https://www.cric.or.jp/english/clj/cl2.html.

odds with this growing trend and could drive massive investments in AI research and supercomputing infrastructure overseas.

Evaluating Whether Outputs Infringe Copyright Requires Attention to Context

The NOI asks whether "AI-generated outputs implicate the exclusive rights of preexisting copyrighted works, such as the right of reproduction or the derivative work right."[53] As an initial matter, as discussed above, OpenAI's products are intended and designed to output new, original expression and also employ a variety of methods to prevent the rare cases where expressive elements of training data might appear in their outputs. But, in the rare situations where an output satisfies copyright law's substantial similarity test, a specific output could implicate the exclusive rights of a copyright owner, depending on the context and facts involved.

In evaluating claims of infringement relating to outputs, the analysis starts with the user. After all, there is no output without a prompt from a user, and the nature of the output is directly influenced by what was asked for. In OpenAI's experience, the overwhelming majority of users have no interest in infringing any exclusive rights in any pre-existing copyrighted work. Experience shows that AI models are not only "capable of substantial noninfringing uses," but are *actually* used *primarily* for entirely non-infringing purposes.[54] Moreover, as described above, OpenAI's models include multiple "guardrails" designed to prevent infringing outputs. Nevertheless, a user deliberately seeking to generate an output that includes those elements may, on occasion, be able to evade those guardrails. Here, again, it is the user who determines whether the output implicates the exclusive rights of a copyright owner. In the argot of copyright law, it is the user who is the "volitional actor."[55]

Furthermore, evaluating whether an output infringes the exclusive rights of a copyright owner turns on how the output is ultimately used, something that is context-specific and falls beyond the control of an AI model developer. For example, as the Supreme Court has recently emphasized, the application of the fair use doctrine often depends on whether an allegedly infringing work is used for a transformative purpose–something that can only be measured by looking at the specific context of the use in question.[56] Accordingly, an output worked up by a competitor "to avoid the drudgery in working up something fresh" would get very different treatment than the same output generated as a critical parody of the original work.[57]

If infringement has occurred, then copyright law's traditional doctrines of secondary liability can be brought to bear to evaluate any potential liability on the part of the AI model creator or service provider. This approach not only reflects the general purpose nature of AI tools and the primacy of the user's contribution, but also accommodates the varied layers of other entities that may be involved. For example, OpenAI makes its models available to other enterprises through its API services. Copyright's secondary liability doctrines are designed to fairly ascribe liability in

---

[53] Question 22, NOI, 88 Fed. Reg. at 59,948.
[54] Sony Corp. of America v. Universal City Studios, Inc., 464 US 417, 442 (1984); see also examples of noninfringing uses cited on pages 1-4, supra.
[55] Perfect 10, Inc. v. Giganews, Inc., 847 F. 3d 657, 666 (2017).
[56] Andy Warhol Foundation Visual Arts v. Goldsmith, 143 S.Ct. 1258, 1276-77 (2023).
[57] Campbell v. Acuff-Rose Music, Inc., 510 US 569, 580 (1994).

these contexts, evaluating issues like the accompanying terms of service, control over user conduct, knowledge of infringing outputs, the intent reflected in marketing materials, and the importance of noninfringing uses.

Copyrightability Questions Can be Addressed by Existing Copyright Laws

The NOI asks several questions about the copyrightability of the outputs of generative AI systems.[58] OpenAI agrees with the Copyright Office that "'statutory text, judicial precedent, and longstanding Copyright Office practice all require human authorship as a condition of copyrightability."[59] OpenAI does not assert any ownership rights over the outputs of its generative AI products.[60]

As the Office recognizes, there remain unanswered questions about when, exactly, sufficient human authorship has been contributed to a work to justify copyright protection. OpenAI's view is that existing copyright law is adequate for addressing these questions as they arise. Given the rapidly evolving nature of generative AI technologies, which in turn are likely to spur rapid changes in the practices of creative industries, it would be premature to attempt to codify a targeted solution at this time.

Moreover, the edge cases of copyrightability are unlikely to be of practical significance in the vast majority of situations. Outputs from ChatGPT and DALL·E, for example, are not likely to be headed directly for commercial exploitation unaltered. Instead, those outputs are typically going to be edited, adapted, and combined with other works before reaching their final form. In those cases, the Office has already confirmed that a work combining human expression with AI-generated material will support a copyright, giving the creator the ability to sell and license the work, and to protect it from wholesale copying.[61]

This may leave some cases where the copyrightability of particular AI-authored elements of a work, standing apart from the other human-authored elements, emerges as an important issue in a future dispute. In those cases, the courts can apply well-established evidentiary and doctrinal tools to separate protected from unprotected elements. The Copyright Office's evolving practices around the registration of AI-assisted works may further assist future fact-finders in addressing these remaining cases as the disputes arise.

---

[58] Questions 18-21, NOI at 59,947-48.

[59] NOI at 59,944 (citing U.S. Copyright Office Review Board, Decision Affirming Refusal of Registration of A Recent Entrance to Paradise at 3 (Feb. 14, 2022), https:// www.copyright.gov/rulings-filings/review-board/docs/a-recent-entrance-to-paradise.pdf).

[60] OpenAI, Terms of Use, https://openai.com/policies/terms-of-use.

[61] U.S. Copyright Office, Cancellation Decision re: Zarya of the Dawn (VAu001480196) at 1 (Feb. 21, 2023), https://www.copyright.gov/docs/zarya-of-the-dawn.pdf.

Conclusion

OpenAI appreciates the opportunity to provide input as part of the NOI process. These are the early days of generative AI systems, and OpenAI urges the Copyright Office to proceed cautiously in calling for new legislative solutions that might prove in hindsight to be premature or misguided as the technology rapidly evolves. As yet, the courts have not had an opportunity to pass on most of the questions raised in the NOI. As discussed above, OpenAI believes that the existing provisions of U.S. copyright law provide a sound foundation on which the courts can build as cases arise.