

Is AI Distillation By DeepSeek IP Theft?

MARCH 12, 2025

This article was originally published in [Law360](#). Reprinted with permission. Any opinions in this article are not those of Winston & Strawn or its clients. The opinions in this article are the authors' opinions only.

The recent unexpected and meteoric rise of an open-source artificial intelligence chatbot developed by DeepSeek, a small Chinese company founded less than two years ago, sent shockwaves throughout the AI and technology communities.

DeepSeek was reportedly developed at a fraction of the cost of the well-known U.S.-based AI chatbots such as OpenAI's ChatGPT, while using one-tenth of the computing power.^[1]

Now, a brewing controversy between OpenAI and DeepSeek has brought the issue of AI distillation and intellectual property rights to the forefront. OpenAI has accused DeepSeek of using a technique called distillation to train its AI models on the outputs of OpenAI's ChatGPT,^[2] raising questions about the legality and ethics of such practices.^[3]

AI DISTILLATION

AI distillation is a process where a smaller, more efficient model (the student) is trained to mimic or distill the outputs of a larger, pretrained model (the teacher). This technique allows the student model to achieve comparable performance to the teacher model while being more cost-effective and easier to deploy.

The process involves the teacher model using real-world data to generate outputs, which are then used as training data for the student model. OpenAI is investigating whether DeepSeek has leveraged the outputs of ChatGPT to train its own models, potentially infringing on OpenAI's IP rights.^[4]

LEGAL IMPLICATIONS

The legal landscape surrounding AI distillation is unclear and evolving. IP law traditionally protects creative works as copyrights, but the application of copyright law to AI-generated outputs is not straightforward.

There is ongoing debate concerning whether the outputs of AI models, such as those generated by ChatGPT, can be subject to copyright protection.^[5] That is, proving a copyright claim in this context could be challenging. One key

issue that remains unclear is whether the outputs of AI models qualify as creative expression, or if they are merely unprotected facts.

Notably, the U.S. Copyright Office published a report on AI and copyrightability on Jan. 29 that sheds light on this issue. The Copyright Office affirmed that current and existing copyright law is enough to address the issues with the new technology.^[6]

The report stated that copyright protection in the U.S. requires human authorship.^[7] Specifically, three scenarios were discussed: (1) prompts, (2) expressive inputs that can be perceived in AI-generated outputs, and (3) modifications or arrangements of AI-generated outputs.

After reviewing the current legal framework and inputs from many commentators, the office concluded that while mere provision of prompts does not render the outputs copyrightable, if a human inputs their own copyrightable work and that work is perceptible in the AI-generated output, or if a human is modifying or arranging AI-generated content, their work would be copyrightable.

In summary, whether an AI-generated product can be copyrighted depends on the amount of human contribution.

Nevertheless, this framework still does not address the specific problem of AI distillation, and it might be hard for OpenAI to argue copyright infringement by DeepSeek.

In OpenAI's own terms of use, it specifically mentions that "[a]s between you and OpenAI ... you (a) retain your ownership rights in Input and (b) own the Output. We hereby assign to you all our right, title, and interest, if any, in and to Output."^[8]

Therefore, even if OpenAI can present enough evidence to show that DeepSeek extracted data from its models, OpenAI likely does not have copyrights over the data. Further legislation may be required to address the specific problem. Additionally, OpenAI's own history of using vast amounts of online data for training its models may provide further defense strategies for DeepSeek.

That is, OpenAI has previously argued that such practices fall under the fair use defense to copyright claims. Specifically, in a 2023 suit, OpenAI faced claims from The New York Times in the U.S. District Court for the Southern District of New York, alleging that its ChatGPT unlawfully used copyrighted content from the newspaper to train their AI systems.

OpenAI argued that its use of the data fell under the fair use doctrine, which allows limited use of copyrighted material without permission for certain purposes such as comment, education or research.^{[9][10]}

This position complicates OpenAI's stance against DeepSeek, as it may be challenging for it to argue that DeepSeek's similar practices constitute IP theft.

ALTERNATIVE WAYS TO ADDRESS AI CONCERNS

Other than copyright, patents and trade secrets could provide alternative recourse against DeepSeek, but those require further investigation into OpenAI's patent portfolio and protected trade secrets.

On the surface, though, these alternative approaches do not currently appear promising. It is unlikely that OpenAI would have patents that cover model outputs alone. Also, OpenAI has not provided specific evidence that DeepSeek illegally accessed its secret data.

As a more promising alternative, non-IP approaches may be considered, such as violation of contracts or user agreements or regulatory measures to restrict the use of DeepSeek in the U.S., e.g., due to concerns with safety of user data. Indeed, several measures are underway.

Open AI has alleged that DeepSeek's practice of distilling the outputs to build rival models is a violation of OpenAI's terms of service. Specifically, in its terms of use, OpenAI states that users cannot "[a]utomatically or programmatically extract data or Output." Although OpenAI has not disclosed what evidence it possesses regarding this alleged violation, using contract law to address AI concerns would provide many companies with better standings than other approaches.

On the public policy front, U.S. agencies such as NASA have swiftly instituted bans against the use of DeepSeek by its employees.^[11] The New York state government has also prohibited their employees from downloading DeepSeek onto state devices.^[12]

Additionally, in February, a bipartisan act, the No DeepSeek on Government Devices Act, was also introduced in the U.S. House of Representatives to ban DeepSeek on all federal employees' government-issued devices.^[13]

CONCLUSION

The dispute between OpenAI and DeepSeek underscores the urgent need for the AI industry to address the legal and ethical challenges posed by AI distillation.

As AI technology continues to advance, it is crucial to establish clear and fair regulations to protect IP while fostering innovation. The development and outcome of this controversy will set important precedents for the future of AI development and IP law.

^[1] J. Vincent, The DeepSeek Panic Reveals an AI World Ready to Blow, The Guardian (Jan. 28, 2025) <https://www.theguardian.com/commentisfree/2025/jan/28/deepseek-r1-ai-world-chinese-chatbot-tech-world-western>.

^[2] <https://www.geeky-gadgets.com/openai-deepseek-intellectual-property-dispute/>.

^[3] <https://www.devdiscourse.com/article/technology/3244686-lutnicks-stand-on-tariffs-and-ai-a-new-era-in-us-commerce>.

^[4] <https://bgr.com/tech/openai-says-it-has-evidence-deepseek-used-chatgpt-to-train-its-ai/>.

^[5] <https://news.yahoo.com/news/openai-little-legal-recourse-against-150858401.html>.

^[6] <https://www.copyright.gov/newsnet/2025/1060.html>.

^[7] <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-2-Copyrightability-Report.pdf>.

^[8] <https://openai.com/policies/row-terms-of-use/>.

^[9] <https://hls.harvard.edu/today/does-chatgpt-violate-new-york-times-copyrights/>.

^[10] <https://www.npr.org/2025/01/14/nx-s1-5258952/new-york-times-openai-microsoft>.

^[11] L. Kolodny, NASA Becomes Latest Federal Agency to Block China's DeepSeek on 'Security and Privacy Concerns,' CNBC (Jan. 31, 2025) <https://www.cnbc.com/2025/01/31/nasa-becomes-latest-federal-agency-to-block-chinas-deepseek.html>.

^[12] <https://abcnews.go.com/US/deepseek-banned-government-devices-new-york-state/story?id=118653885>.

^[13] <https://gottheimer.house.gov/posts/release-gottheimer-lahood-introduce-new-bipartisan-legislation-to-protect-americans-from-deepseek>.

Related Capabilities

Intellectual Property

Patent Litigation

Related Professionals



Brian E. Ferguson



Chaoxuan Charles Liu



Yuchen Han