



June 4, 2024

Via email

Suzanne V. Wilson, General Counsel and
Associate Register of Copyrights
United States Copyright Office
101 Independence Ave. S.E.
Washington, DC 20559

Re: Summary of *Ex Parte* Meeting Regarding Copyright Office Artificial Intelligence Study, Docket 2023–6

Dear Ms. Wilson,

Pursuant to 37 C.F.R. § 205.24(c)(3), we submit this letter summarizing our 30-minute meeting on May 28, 2024 in connection with the Office’s Artificial Intelligence (“AI”) Study.

The OpenAI participants at the meeting were Tom Rubin, Chief of Intellectual Property and Content, and Chan Park, Head of U.S. and Canada Policy and Partnerships.

The Copyright Office participants at the meeting were:

- Suzanne Wilson – General Counsel and Associate Register of Copyrights
- Emily Chapuis – Deputy General Counsel
- Andrew Foglia – Deputy Director for Policy and International Affairs
- Nick Bartelt – Assistant General Counsel
- Brandy Karl – Assistant General Counsel
- Jason Sloan – Assistant General Counsel
- Gabi Luna – Paralegal Specialist
- Michael Druckman – Attorney-Advisor
- Isaac Klipstein – Barbara A. Ringer Fellow
- Chris Weston – Senior Counsel for Policy and International Affairs
- Caitlin Costello – Counsel for Policy and International Affairs
- Jenée Iyer – Counsel for Policy and International Affairs
- Danielle Johnson – Counsel for Policy and International Affairs
- Aaron Watson – Attorney-Advisor

We covered two topics at our meeting: 1) data sourcing, and 2) digital replicas.

Data Sourcing

We opened by discussing our recent announcement that we are building a product called Media Manager, which will allow creators and copyright owners to tell us “what they own and specify how they want their works to be included or excluded from machine learning research and training.”¹ We explained that Media Manager will serve as a granular way for copyright owners to indicate use preferences for specific works, regardless of media type or where the works may be present on the internet. We are designing Media Manager to identify works based on how they appear or are otherwise perceived, rather than expecting works to carry metadata that could be removed. We plan to launch Media Manager in early 2025 and offer it to copyright owners globally.

We further explained that the initial version of Media Manager will be focused on machine learning research and training, but we plan to continue adding to the product after launch. For example, we are exploring whether Media Manager could be made interoperable with other AI actors so that OpenAI can respect preferences shared with other organizations and they can do the same with Media Manager. We are also considering whether Media Manager could be expanded to other forms of intellectual property, such as to address digital impersonation concerns. But our first priority is designing the product in consultation with creators and regulators so that it meets their immediate needs.

We also shared additional specifics about how we use publicly available data. In addition to respecting machine-readable preferences on websites through robots.txt, we affirmatively avoid crawling websites that we know employ paywalls for access to their content, websites that primarily aggregate personally identifiable information, and websites that have content that violates our policies.² For example, we exclude sites that are known to engage in IP infringement, such as the “notorious markets” identified annually by the Office of the U.S. Trade Representative.³ We also exclude domains that have been flagged to us by copyright owners and their representatives as hosting infringing content.

Finally, we discussed our approach to content partnerships. While we do not enter partnerships solely for publicly available data, we do partner for non-public data (such as archival data and metadata) and to use copyrighted material in our products in ways that would exceed fair use, such as displaying long excerpts from a work in response to a user

¹ See *Our approach to data and AI*, OpenAI (May 7, 2024), <https://openai.com/index/approach-to-data-and-ai/>.

² *Id.*

³ See *2023 Review of Notorious Markets for Counterfeiting and Piracy*, Office of the U.S. Trade Representative (2024), https://ustr.gov/sites/default/files/2023_Review_of_Notorious_Markets_for_Counterfeiting_and_Piracy_Notorious_Markets_List_final.pdf.

query.⁴ We expect these partnerships will become more common in the industry as AI providers identify new ways that copyrighted material can benefit users and provide value back to publishers. But we also explained to the Office that, given the quantity and breadth of data required to train advanced AI models, it is not feasible to source the necessary quantity of data from partnerships.

Digital Replicas

We then shared our perspective on digital impersonation, which is informed by our experience building a synthetic voice model called Voice Engine. As we recently announced, we have built a text-to-speech model—Voice Engine—that can generate natural-sounding audio from a single 15-second audio clip of a target voice.⁵ Voice Engine is not publicly available, but a few trusted partners are testing the model to help us understand beneficial applications of the technology⁶ and identify potential mitigations for harm.

We also explained our protections against misuse. Currently, we require partners to disclose that their voices were generated by AI, require partners to obtain explicit and informed consent from original speakers, and use watermarking techniques so that we can identify and monitor where Voice Engine output appears.⁷ We are also testing additional measures, including requiring an input voice to read randomly generated audio to prove the speaker is present and aware of how their voice is being used.⁸

Finally, we shared how Voice Engine has informed our perspective on the need for legal protections against deceptive digital impersonation. Given that it is possible as a technical matter to generate convincing synthetic audio, we support the creation of legal protections against AI-powered impersonation. We believe that AI models should be developed while considering potential harms to content creators and the general public, and with respect for social norms.⁹ The use of AI to create deceptive digital replicas does not align with those values.

⁴ *Our approach to data and AI*, OpenAI (May 7, 2024), <https://openai.com/index/approach-to-data-and-ai/> (explaining that OpenAI “partner[s] to access non-publicly available content, such as archives and metadata”).

⁵ See *Navigating the Challenges and Opportunities of Synthetic Voices*, OpenAI (Mar. 29, 2024), <https://openai.com/index/navigating-the-challenges-and-opportunities-of-synthetic-voices/>.

⁶ Among other examples, our partners have employed Voice Engine to generate translated videos and help patients with generative speech conditions recover their voice. See *id.* (discussing work by HeyGen, Lifespan, and others).

⁷ *Id.*

⁸ See Kyle Wiggers, *OpenAI built a voice cloning tool, but you can’t use it...yet*, TechCrunch (Mar. 29, 2024), <https://techcrunch.com/2024/03/29/openai-custom-voice-engine-preview>.

⁹ See *Introducing the Model Spec*, OpenAI (May 8, 2024), <https://openai.com/index/introducing-the-model-spec/> (setting out initial draft of OpenAI’s objectives and rules for shaping model behavior).

While we did not offer a specific legal proposal, we recommended that the Office carefully consider the role of service providers in such a regime. Like many technologies, synthetic media can have both beneficial and harmful impacts depending on how it is used. Service providers should be able to offer positive, consensual uses of synthetic media without being held responsible for those who would abuse the technology. We encouraged the Office to examine whether a form of safe harbor would be appropriate for technology providers that do not induce users to create non-consensual digital replicas and take proactive steps to monitor and mitigate harmful uses.

* * *

We thank the Copyright Office for taking the time to meet with us and discuss these issues. Please do not hesitate to contact me if you have any further questions.

Sincerely,

Tom Rubin
Chief of Intellectual Property and Content
OpenAI