UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

IN RE: OPENAI, INC., COPYRIGHT INFRINGEMENT LITIGATION This Document Relates To: MICHAEL CHABON v. OPENAI, INC., et al., No. 1:25-cv-03291-SHS-OTW PAUL TREMBLAY V. OPENAI, INC., et al., No. 1:25-cv-03482-SHS-OTW SARAH SILVERMAN V. OPENAI, INC., et al. No. 1:25-cv-03483-SHS-OTW NICHOLAS A. BASBANES v. MICROSOFT CORPORATION, et al., No. 1:24-cv-00084-SHS-OTW AUTHORS GUILD, et al. v. OPENAI, INC., et al., No. 1:23-cv-08292-SHS-OTW JONATHAN ALTER V. OPENAI, INC., et al., No. 1:23-cv-10211-SHS-OTW	25-md-3143 (SHS) (OTW) 25-cv-03291-SHS-OTW 25-cv-03482-SHS-OTW 25-cv-03483-SHS-OTW 24-cv-00084-SHS-OTW 23-cv-08292-SHS-OTW 23-cv-10211-SHS-OTW Hon. Sidney H. Stein Hon. Ona T. Wang

DECLARATION OF BRIGHT KELLOGG

I, Bright Kellogg, hereby declare as follows:

1. I have been an employee of OpenAI OpCo, LLC ("OpenAI") since 2024. I am currently a Program Manager and member of the Legal Operations staff. In that role, I manage projects and programs related to litigation/regulatory needs. Through my employment at OpenAI, I am generally familiar with the training process for OpenAI's Large Language Models ("LLMs") as well as OpenAI's LLM development history. I am also familiar with the process for setting up and hosting OpenAI's LLMs and making them accessible via user queries to OpenAI's Application Programming Interface ("API").

2. I have knowledge of the facts set forth below based upon personal experience or based upon information from other OpenAI personnel who have personal experience and if called as a witness, could testify competently thereto, to the best of my knowledge and belief.

3. The process of training a model that OpenAI makes available as part of a commercial product (a "production model") is complex and results in hundreds or even thousands of intermediate versions of the model that are, in effect, incomplete "drafts" of the production model.

4. Many incomplete draft models are created throughout the training process (described below), and even more are created to implement continuous model upgrades to address issues such as bug fixes or performance enhancements. These incomplete draft models are never put into production and never made available for public use.

5. To the extent OpenAI stores these incomplete draft models, it does so to assist in model development, *e.g.*, because they can be used as "checkpoints" during the training process so that researchers can return to a previous checkpoint in the event a particular upgrade or training run does not work or for purposes of developing additional models. OpenAI does not

comprehensively catalogue these incomplete draft models or store them in a central repository, nor does it maintain a central store of information regarding the function or purpose of each incomplete draft model within the broader model development process.

6. Locating a specific draft model—and then determining the significance of any given checkpoint—may require a burdensome forensic search by OpenAI's technical staff that could take weeks if not months for each draft model, including document investigation and review as well as individual interviews. That process would be even more challenging for older models like GPT-3.5, GPT-3.5 Turbo, GPT-4, or GPT-4 Turbo, which were created in large part by employees who are no longer at the company.

7. One part of the model development process is called "pre-training," which is an initial learning phase during which OpenAI uses a vast collection of data to identify linguistic patterns, common facts, and concepts that underlie human communication. The draft models that result from this initial learning phase are incomplete: they have not been trained to understand and respond to user questions, nor have they been trained to adhere to OpenAI's safety standards and behavior policy. They are also internal: they are not intended to be made available for use and are not used to power any of OpenAI's products.

8. Instead, these incomplete draft models must undertake a variety of further steps before they can be used in an external product. Some of these steps are collectively referred to as "post-training." Through these additional training steps, OpenAI shapes the models' behavior to make them useful, including by training the models to understand and respond to user questions, to generate responses that are useful and safe, and to improve the models' ability to understand and generate long sequences of text.

9. During "post-training," a model might undergo dozens or even hundreds of individual "post-training" runs. Each step in this process results in yet another incomplete draft of the model. As a result, multiple drafts of a "post-trained" model exist.

10. During development of its GPT-class models and prior to their release, OpenAI has also provided select external researchers and red-teamers with access to intermediate drafts of its model. This would include, for example, the two specific draft models referred to publicly as "GPT-3.5 base" and "GPT-4.0 base." These intermediate draft models were pre-trained; some draft models may have undergone some additional post-training. During this testing process, the externally available version of the model was often continuously updated. There was no single "checkpoint" offered externally, and there is no central log recording what changes each version of the checkpoint reflects.

11. Once OpenAI completes the training process, it must conduct additional technical work to make that model accessible through one of its products, including through its application programming interface or "API." OpenAI does not typically undertake this process for incomplete draft models.

12. OpenAI does not have an established process or infrastructure to expose older draft checkpoints through a user-facing API. Making draft models available through OpenAI's API would be a complicated and burdensome technical exercise requiring approximately six calendar weeks of cross-functional engineering effort per draft model. OpenAI would have to write code to convert and package a draft model to ensure its compatibility with OpenAI's API engine and schema. In addition, for each model, OpenAI would also have to build a one-off, access-restricted path into its model registry and deployment pipelines; provision and validate dedicated serving infrastructure and capacity; and implement secure telemetry, audit logging, billing, and access

control hooks prior to any limited exposure of the draft model through OpenAI's API. That a draft checkpoint was previously made available does not necessarily alleviate this burden, because OpenAI does not expect these checkpoints to be restored and, as such, would need to reconfigure draft checkpoints to make them accessible.

13. Drafts of certain older models, such as GPT-3.5, GPT-3.5 Turbo, GPT-4, and GPT-4 Turbo, are likely incompatible with OpenAI's current "inference engine," which executes OpenAI's models to generate responses based on user queries. Building a new inference engine that is compatible with older draft models would require additional, bespoke engineering effort on top of the estimated six calendar weeks per model.

14. Once a model has been engineered so that it can be made accessible through OpenAI's API, the model must then be deployed and hosted on an allocation of graphics processing units (GPUs). Deploying and hosting deprecated or never-deployed models, if even possible, would require potentially weeks of additional engineering effort.

I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct.

Dated: November 21, 2025

BRIGHT KELLOGG