

OpenAI Response to the European Commission Call for Contributions “*Competition in Generative AI*”¹

OpenAI welcomes the opportunity to respond to the Commission’s call for contributions on competition in generative AI systems. This contribution explains OpenAI’s current views on the inputs required to succeed in the emerging generative AI marketplace, the challenges facing the industry, and that the current AI ecosystem is vibrant and rapidly evolving.

As the generative AI industry evolves rapidly, an increasing number of companies, laboratories, and other public and private institutions are developing and deploying or otherwise making available an incredible diversity of AI models and features, which millions of users are using or building upon every week. The industry’s continued health and success depends on actors’ ability and incentive to invest in research and deployment, which requires funding that supports broad and reliable access to compute, data, and talent.

While there are areas in which competition enforcers need to be vigilant, as discussed below, existing competition rules and approaches appear flexible and capable of protecting and promoting growth and innovation in generative AI.

The following sections address the Commission’s questions.

1. What are the main components (i.e., inputs) necessary to build, train, deploy and distribute generative AI systems? Please explain the importance of these components.

&

7. What is the role of data and what are its relevant characteristics for the provision of generative AI systems and/or components, including AI models?

Building, training, deploying, and distributing generative AI systems like those offered by OpenAI requires several key components, most notably: (1) sufficient funding; (2) engineering talent; (3) significant computing resources, i.e. graphics processing units (GPUs) and cloud capacity; and (4) large and diverse data.

- **Funding.** Generative AI can be resource intensive, particularly for the training and deployment of the most highly and broadly capable models. Initial funding and sustained investment are important for securing talent and compute and could become increasingly important for data. Recent public reports reflect the critical

¹ European Commission, *Competition in Virtual Worlds and Generative AI, Calls for Contributions* (Jan. 9, 2024), https://competition-policy.ec.europa.eu/system/files/2024-01/20240109_call-for-contributions_virtual-worlds_and_generative-AI.pdf [herein “EC Generative AI Call for Contributions”]. References to “generative AI” include the “generative AI systems and/or components, including AI models,” addressed in the EC Generative AI Call for Contributions.

role that funding plays across the industry. In 2023, generative AI and AI-related startups raised nearly \$50 billion—up 9% from the \$45.8 billion invested in the sector in 2022.² [REDACTED]

[REDACTED] OpenAI must raise capital and generate subscription, usage, or other revenue, to continue to fund its research, development and deployment of models to support its mission of developing safe artificial general intelligence ("AGI") for the benefit of humanity. [REDACTED]

- **Talent.** Generative AI requires highly trained talent. Data scientists, machine learning engineers, and other researchers develop training processes, design architecture, optimize speed and performance and fine-tune for usefulness, accuracy, and safety. This highly skilled workforce is mobile and in high demand, allowing the AI talent pool the flexibility to transfer to, and to create a great number and variety of AI-related companies worldwide. OpenAI competes hard for talent with other AI companies of all sizes, including start-ups and much larger players, academia, and other research institutions, each of which provides scientific and/or commercial opportunities to develop next-generation products and services across the generative AI stack. However, unlike in some areas of cutting-edge technology development, you do not need thousands of engineers or data scientists. Many companies developing and deploying foundation models, including OpenAI, have tens or hundreds of engineers and AI researchers working on foundation models, not thousands or tens of thousands. This means many AI companies using small cadres of skilled engineers and data scientists are innovating, developing, and deploying very competitive AI models and AI products and services.
- **Compute.** Generative AI models require significant computational resources ("compute"). GPUs, or similar advanced chips, are a critical input for the

² Chris Metinko, *Artificial Buildup: AI Startups Were Hot In 2023, But This Year May Be Slightly Different*, Crunchbase News (Jan. 9, 2024), <https://news.crunchbase.com/ai/hot-startups-2023-openai-anthropic-forecast-2024/> ("Generative AI and AI-related startups raised nearly \$50 billion in 2023, per Crunchbase data, with some, including OpenAI, Anthropic and Inflection AI, raking in billions of dollars all by themselves."); Gene Teare, *Global Startup Funding In 2023 Closes In At Lowest Level In 5 Years*, *Crunchbase News* (Jan. 4, 2024), <https://news.crunchbase.com/venture/global-funding-data-analysis-ai-eoy-2023/> ("While most industries were down year over year, AI was the largest sector to show an increase. Global funding to AI startups reached close to \$50 billion last year, up 9% from the \$45.8 billion invested in 2022.").

development of generative AI due to their performance of multiple computational operations in parallel. Compute is a scarce and expensive resource that requires ongoing and increasing investment. So far, this scarcity and cost does not appear to be preventing a broad range of companies of different sizes and focus areas from innovating and competing across generative AI models, products, and services.

[REDACTED]

- **Data.** Generative AI systems require diverse and very large data sources. The continued availability of public datasets is crucial for pre-training generative AI models because they are accessible to a wide range of researchers and developers, which promotes competition and innovation. Many prominent foundation models have been pre-trained entirely on public sources, including LLaMA (Meta) and Stable Diffusion (StabilityAI). For continued evolution of models, significant additional data and more diverse data (across many aspects including source, language, and medium (e.g., text, images, audio, and video)) will be important.

Both the volume and diversity of data are particularly important for pre-training a model. For example, a large-language model (“LLM”) is exposed to a huge variety of text from which it will learn the fundamentals of language like the relationships between words and the rules of grammar, syntax, and usage. Similarly, image or video generators work best when they learn from a large volume and diversity of images or videos. We note that despite a common and unfortunate misperception of the technology, 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. 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

³ [REDACTED]

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.

[REDACTED]

Pre-training results in a “base” model that is very capable at predicting the next word in a sequence, including in novel sequences. But a base model is not necessarily useful or safe. “Post-training” denotes the additional training and other engineering work necessary to turn a base model into a safer, useful model. Because the goal of this “post-training” is generally to teach a model different behavior and align it with human preferences (rather than fundamental language understanding), post-training data is often much more specialized. For example, post-training often relies on relatively small, high-quality datasets that establish a standard of ideal model behavior. Using a technique known as reinforcement learning from human feedback (RLHF), these ideal answers teach the model to (1) follow instructions, (2) decrease the likelihood of returning inaccurate content, and (3) add safety features that check the model’s output.

Models can also rely on specialized data for fine-tuning to increase performance in particular areas. For example, a customer may begin with an open-source model or a proprietary model, and then fine-tune the model on an *internal customer* dataset. Using the internal customer dataset can, for example, improve the model’s ability to provide more customized answers to customer service queries or create more helpful suggestions based on past data in that company’s knowledge repositories.

2. What are the main barriers to entry and expansion for the provision, distribution or integration of generative AI systems and/or components, including AI models? Please indicate to which components they relate.

&

4. Which competition issues will likely emerge for the provision, distribution, or integration of generative AI systems and/or components, including AI models. Please indicate to which components they relate.

The generative AI ecosystem is vibrant and growing, with access to compute and continued access to new data sources for pre-training being the biggest challenges. OpenAI does not believe that access to data or distribution at this point poses a significant barrier to new entrants or expansion of existing companies. [REDACTED]

[REDACTED]

As noted above, funding is a critical input to the provision of generative AI. When large companies fund startups and less established companies, they fuel innovation and provide expertise and capital needed for other critical AI inputs—most notably compute and talent.

[REDACTED]

More global investment in compute development and manufacturing would help expand and diversify the supply of chips for AI training and inference. Significantly expanding supply of the most advanced chips will be important for continued competitive development and broad adoption of AI systems, increasing innovation and fueling a more robust foundation model ecosystem.

Meeting growing future compute needs and building the infrastructure needed to operate that compute, will require not only additional private investment, but also government investment and incentives for private investment and public-private partnerships. Among the laudable public efforts is the European Union’s initiative to boost the supply of computational resources, The European High Performance Computing Joint Undertaking (“EuroHPC JU”).⁵ The EuroHPC JU aims to assist in the growth and adoption of innovative and competitive supercomputing systems and to expand the access of the supercomputing infrastructure to an extensive range of users

⁴ [REDACTED]

⁵ European Commission, The European High Performance Computing Joint Undertaking, <https://digital-strategy.ec.europa.eu/en/policies/high-performance-computing-joint-undertaking>.

from both the public and private sectors.⁶ OpenAI similarly supports the CHIPS and Science Act in the United States, which represents a significant investment in U.S. semiconductor manufacturing, research and development, and workforce enhancement, totaling nearly \$53 billion.⁷ Through a combination of public and private investment, the U.S. aims to significantly increase its production of leading-edge chips by 2030.⁸

[REDACTED]

In sum, to protect against barriers to entry, OpenAI supports: (1) recognition of the importance of partnerships between large and small entities to ensure necessary funding for research, talent, and compute; (2) ongoing public and private investment in compute infrastructure; (3) provisions to protect access to diverse data sources necessary for training; and (4) careful antitrust enforcement—using existing tools—against foreclosure of access to smaller and/or nascent entrants from the marketplace.

3. What are the main drivers of competition (i.e. the elements that make a company a successful player) for the provision, distribution or integration of generative AI systems and/or components including AI models?

The environment for generative AI systems is competitive and quickly evolving, marked by significant investments and rapid technological advancements by a wide range of players. Users can choose from an array of AI models to meet their needs and are choosing AI models that are fit-for-purpose, and often choosing multiple models for different needs. Among the options from which users can choose, mix, and match are proprietary and open-source models designed for a wide array of use cases, with more models and new features of existing models being introduced every day. Many

⁶ Discover EuroHPC JU, https://eurohpc-ju.europa.eu/about/discover-eurohpc-ju_en.

⁷ Press Release, The White House, United States of America, *FACT SHEET: One Year after the CHIPS and Science Act, Biden-Harris Administration Marks Historic Progress in Bringing Semiconductor Supply Chains Home, Supporting Innovation, and Protecting National Security* (Aug. 9, 2023), <https://www.whitehouse.gov/briefing-room/statements-releases/2023/08/09/fact-sheet-one-year-after-the-chips-and-science-act-biden-harris-administration-marks-historic-progress-in-bringing-semiconductor-supply-chains-home-supporting-innovation-and-protecting-national-s/>.

⁸ Remarks by U.S. Secretary of Commerce Gina Raimondo: Investing in Leading-Edge Technology: An Update on CHIPS Act Implementation (Feb. 26, 2024), <https://www.commerce.gov/news/speeches/2024/02/remarks-us-secretary-commerce-gina-raimondo-investing-leading-edge-technology>.

companies, labs, and other institutions offer models with different performance, capabilities, and price points, as well as specialized models such as those for image generation or text-to-speech. This dynamic and diversified environment creates a highly competitive marketplace for generative AI, that has seen companies reduce pricing and constantly innovate across capabilities, features, interface and many other aspects to attract and keep users.

[REDACTED]

Moreover, the generative AI sector is intrinsically more competitive than other technology sectors such as semiconductors, which have an extremely complex multi-stage supply chain involving thousands of parallel innovations coordinated across potentially thousands of engineers. In contrast, developing and training AI models (assuming sufficient chips and data) may require only dozens or hundreds of researchers and engineers. OpenAI is relatively small [REDACTED] and faces fierce competition from even smaller companies, as well as much larger ones.

Further, research by OpenAI and others has shown that there are often many ways to train a language model. Algorithmic differences that matter at small scale may end up being insignificant at large scale. For a new generative AI developer to achieve similar performance to the leading players, particularly in select areas of performance, it doesn't necessarily need equivalent amounts of data and compute. In generative AI it is possible to launch and develop a successful product with modest amounts of those inputs and then to scale up, particularly if initially building on an open-source or existing proprietary model.

5. How will generative AI systems and/or components, including AI models, likely be monetized, and which components will likely capture most of this monetization?

Generative AI is different than other technology services offered on the internet or through software where marginal cost is very low. For AI models, the cost of compute means the marginal cost of running an AI model is significant. To meet this challenge, AI companies appear to be using several revenue models, including subscription-based, usage-based, project based, and advertising-supported approaches.

[REDACTED]

[REDACTED]

Because the AI space is very competitive, API pricing and chat usage have experienced steady price decreases (API) and/or value, quality, and feature increases (chat options). This means users are able to choose among services at different and very competitive price points that have continued to decrease while features and capabilities increase.

The breadth of AI companies and AI use cases also likely means that the variety of approaches to monetizing AI will continue.

6. Do open-source generative AI systems and/or components, including AI models, compete effectively with proprietary AI generative systems and/or components? Please elaborate on your answer.

Many open-source AI models, including Mistral's Mixtral, Meta's Llama2, Google's Gemma, as well as many others, including some of the thousands on Hugging Face, compete very effectively with proprietary AI systems. Companies and developers are building off open-source models and fine-tuned open-source systems can have very competitive capabilities and performance against many proprietary models.

[REDACTED]

Moreover, many developers have built their own proprietary models based on open-source models, on which they are offering model access, as well as new AI products and services. For consumers there are a plethora of new AI products and service offerings, with some based on proprietary models and others built off open-source models.

OpenAI has made several of its models available on an open-source basis, including Whisper v3 released in November 2023. We also open source tools that are designed to help make all models (not just our models) safer, such as our suite of OpenAI evals.

8. What is the role of interoperability in the provision of generative AI systems and/or components, including AI models? Is the lack of interoperability between components a risk to effective competition?

In the context of generative AI systems and/or components, including AI models, it is unclear what interoperability means or how interoperability would be workable or beneficial for competition. Interoperability makes sense in the telecommunications industry because those standards allow users to communicate with users on other systems—a core function of telecommunications. However, requiring AI systems to interoperate would seem to require models to operate very similarly, be trained on the same data in the same way, have the same post-training, and the same user interfaces, which is very different from how the AI ecosystem operates today. Artificial or intentional barriers to interoperability should be discouraged, but at least to date, that doesn't seem to be an issue. At this point, interoperability requirements would seem to be solving for a problem that is not apparent and would risk hindering competitiveness and innovation in the industry.

9. Do the vertically integrated companies, which provide several components along the value chain of generative AI systems (including user facing applications and plug-ins), enjoy an advantage compared to other companies? Please elaborate on your answer.

[REDACTED]

[REDACTED]

[REDACTED]

10. What is the rationale of the investments and/or acquisitions of large companies in small providers of generative AI systems and/or components, including AI models? How will they affect competition?

As noted above, reliable and adequate funding has been critical to the growth of generative AI. Investments help generative AI companies hire and retain key talent and acquire or access compute and cloud services, without having to build their own data centers or have the expertise to build their own supercomputers. For larger companies investments in smaller companies can provide insights into generative AI developments, access to and collaborations on IP, and help drive additional traffic and earnings on their clouds or greater demand for their chips or other services. The number of large companies investing in many different smaller providers of generative AI systems, models, and other components has helped fuel the growth and expansion of the industry, and innovation that would have taken longer and been less widely available without these investments.

For OpenAI, Microsoft’s funding that started in 2019 and increased in 2021 and 2023, has been essential for OpenAI to sustain its research and to develop, test, and deploy new models to a broad swath of customers (consumers and businesses) including on a free basis to millions of users, and to further its mission of developing safe AGI for the benefit of humanity.

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

While OpenAI has a unique structure, many other companies have entered into partnerships with larger, more established companies with greater compute and data center resources as well as very significant capital holdings. For example, Anthropic works with AWS and Google; A121 partners with Google; Inflection.AI works with Nvidia and Microsoft; Mistral partners with Microsoft and Google; Databricks and TogetherAI both work with Nvidia. This is only a sampling of existing partnerships. Each of these partnerships between smaller and larger firms reflects the important role that funding plays in launching and sustaining innovation and competition in generative AI, as well as providing additional expertise and access to compute and cloud services.

11. & 12. Do you expect the emergence of generative AI systems and/or components, including AI models to trigger the need to adapt EU legal antitrust concepts? Do you expect the emergence of generative AI systems to trigger the need to adapt EU antitrust investigation tools and practices?

EU antitrust concepts are robust, flexible, and broadly applicable across the AI industry. As discussed above, generative AI currently is evolving in a competitive environment characterized by innovation, entry, and expansion, at many different levels and across companies throughout the worldwide ecosystem, with both proprietary and open-source models driving growth, diversity, usage, and innovation. To the extent antitrust issues arise—[REDACTED]—existing EU laws and tools appear equipped to address such concerns. At this early stage in the AI industry's development—[REDACTED]—with a broad and complex ecosystem of competing models, and a wide range of system level innovations, it is unclear what new rules, tools, or practices beyond those already existing and available to enforcers would be needed or beneficial.