

Non-Confidential Version

[Redacted]’s Response to the European Commission’s Call for Contributions on Generative Artificial Intelligence

April 16, 2024

[Redacted] appreciates the opportunity to provide information regarding the generative artificial intelligence (“gen AI”) sector. [Redacted].¹

Our submission covers the following issues: (1) the need for gen AI development to be consistent with the European Digital Markets Act (DMA); (2) the need for gen AI development to avoid creating or entrenching walled gardens; (3) NVIDIA’s position in AI chips and CUDA’s role; and (4) the importance of access to data that is open and consistent with all applicable laws.

I. Gen AI Development Should be Consistent with DMA

While DMA is unlikely to be applicable to all areas of gen AI development, it may apply to certain aspects of gen AI implementation, such as the role of operating systems in the deployment of gen AI models, developers’ access to data (which can be used for training, fine-tuning, or running gen AI models), or the integration of gen AI into core platform services such as online intermediation services, AI Agents/co-pilots or search engines. Where applicable, gatekeepers need to comply with their obligations under DMA with respect to the development and deployment of gen AI. DMA coming into force just as the ecosystem for gen AI begins to flourish is an important factor in enabling an open and competitive AI ecosystem, at least in relation to the digital markets subject to DMA.

This need for compliance with DMA does not mean that DMA alone can address all competition issues arising from the development and implementation of gen AI. First, several other aspects of gen AI development and implementation are likely not covered by DMA, such as the supply of infrastructure for gen AI development. Second, DMA’s definition of gatekeepers does not apply to all instances in which a stakeholder may have monopoly power in a well-defined market. Third, DMA does not address all types of conduct that may give rise to anticompetitive effect. These considerations are consistent with the DMA’s role as an ex-ante regulatory scheme for certain digital markets separate from competition law, but it means that the development and implementation of gen AI, whether subject to DMA or not, must also comply with competition laws in all applicable jurisdictions.

¹ 7B denotes 7 billion parameters. 13B denotes 13 billion parameters.

II. Gen AI Development and Implementation Should Not Create or Entrench Walled Gardens

Consistent with the point above, gen AI development should not enable stakeholders to entrench their walled gardens or create new ones. For example, gen AI development should not make it more difficult for developers and users to switch from iOS to Android and vice versa. Conversely, agencies should be vigilant against the risk of gen AI being developed into closed ecosystems, including through the leveraging of positions in other markets.

The issue of high entry barriers to certain digital markets caused by, among other things, high switching costs and data dependency, is well-recognized.² In fact, the European Commission explained that it is one of the driving factors leading to the problem that DMA was designed to address.³ Because several digital products and services are involved in the development and deployment of gen AI and conversely are affected by gen AI implementations, gen AI development may exacerbate the entrenchment of digital walled gardens. On the other hand, certain firms may be able to leverage their position in adjacent digital markets to build closed gen AI ecosystems.

At this early stage of gen AI development and implementation, it is not possible to foresee all the ways in which gen AI may create or entrench walled gardens. Even when market participants do not engage in anticompetitive conduct, a market susceptible to tipping may still eventually lose competitiveness and become a winner-takes-all market.⁴ Depending on how gen AI continues to be developed and deployed, it may turn out to be one such market.⁵ Therefore, to prevent a possibly irreversible walling off of ecosystems in the future, agencies should begin to examine the development and deployment of gen AI and its potential effects on competition.

III. NVIDIA Has High Share of Chips for Gen AI and CUDA Contributes to This Position

² See, e.g. European Commission, *Impact Assessment Report Accompanying Proposal for a Regulation of the European Parliament and of the Council on Contestable and Fair Markets in the Digital Sector (Digital Markets Act)*, SWD(2020) 363, ¶¶ 69, 73 (Dec. 15, 2020).

³ *Id.*

⁴ See *id.* ¶ 120 (“Second, market failures associated to tipping markets cannot be tackled on the basis of the existing competition rules, notably where market tipping is triggered primarily by the market structure, and not (or only to a lesser extent) by any specific conduct.”).

⁵ See, e.g., U.S. Federal Trade Commission Staff in the Bureau of Competition & Office of Technology, *Generative AI Raises Competition Concerns*, Technology Blog (Jun 28, 2023) (“FTC Gen AI Blog Post”), <https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns> (“Firms in generative AI markets could take advantage of network effects to maintain a dominant position or concentrate market power . . . Absent legal or policy intervention, network effects can supercharge a company’s ability and incentive to engage in unfair methods of competition.”).

Access to computational resources, including chips, is reported to be a “key input” to gen AI markets.⁶ Market researchers report that NVIDIA commands significant pricing power for its GPUs for AI usage.⁷ Various analysts estimate NVIDIA’s share to be over 80% depending on the definition of the subsegments of AI chips.⁸

To assist the Commission in its inquiry, [redacted].

Analysts have explained that NVIDIA’s proprietary CUDA framework contributes to NVIDIA’s position in AI chips.⁹ Specifically, NVIDIA first introduced the CUDA framework in 2006 to enable developers to write software on its GPUs, especially for parallel computing (which will later be used in gen AI development). Due to the popularity of NVIDIA’s GPUs, CUDA also became widely used by developers using the GPUs, including for gen AI training. If a developer wants to use a different framework, they need to rewrite the entire software or a significant part of it. Because of the existing developer base on CUDA and the associated available resources, developers writing new software have an incentive to use CUDA rather than alternative frameworks. Further, with resource-intensive use cases like gen AI training, developers are unlikely to have the time or resource to develop different software or gen AI model on an alternative framework.

In addition, once a foundation model has been trained on NVIDIA GPUs and CUDA, it is very difficult to adapt the model to run on different hardware at the edge. The adaptation process for the model to run on different hardware takes intensive effort and significant time. This is likely to decrease developers’ incentives to deploy gen AI in edge devices. Further, in the PC segment, developers may have an incentive to use NVIDIA’s PC chips for gen AI applications due to higher compatibility between the

⁶ *Id.*

⁷ See, e.g., The Motley Fool, *2 Artificial Intelligence (AI) Stocks That Could Help Make You a Fortune* (Jan. 23, 2024), https://www.fool.com/investing/2024/01/23/2-artificial-intelligence-stocks-that-could-help/?utm_source=nasdaq&utm_medium=feed&utm_campaign=article&referring_guid=4170d96f-ea5f-4419-8dd7-12e07cf06682 (“For example, Nvidia’s flagship H100 AI graphics processing unit (GPU) -- which is in hot demand and reportedly commands a waiting period between 36 and 52 weeks -- reportedly costs \$3,320 to manufacture, according to investment bank Raymond James. Nvidia reportedly sells each H100 processor at an average price of \$30,000, commanding a huge profit on each unit of this AI chip.”).

⁸ See, e.g., Extreme Tech, *Analysts Estimate Nvidia Owns 98% of the Data Center GPU Market* (Feb. 1, 2024), <https://www.extremetech.com/computing/analysts-estimate-nvidia-owns-98-of-the-data-center-gpu-market> (“[A global shortage of NVIDIA GPUs for AI] has reportedly led Nvidia to capture 98% of the current data center GPU market, according to analysts from Wells Fargo.”); Reuters, *With no big customers named, AMD’s AI chip challenge to Nvidia remains uphill fight* (Jun 13, 2023), <https://www.reuters.com/technology/amd-likely-offer-details-ai-chip-challenge-nvidia-2023-06-13/> (“Nvidia . . . dominates the AI computing market with a market share of 80% to 95%, according to analysts.”). See also U.S. Federal Trade Commission, *Tech Summit on Artificial Intelligence: A Quote Book, Hardware and Infrastructure Edition: Semiconductor Chips & Cloud Services*, 9 (Mar. 2024), https://www.ftc.gov/system/files/ftc_gov/pdf/ftcotstaffreport-techaisummitquotebook.pdf (quoting Corey Quinn, Chief Cloud Economist, The Duckbill Group, statement that “the biggest challenge that we’re seeing is that all roads lead to Nvidia. They are today a bottleneck on all of this . . .”).

⁹ See, e.g., Dan Gallagher, *How Nvidia Got Huge—and Almost Invincible*, *The Wall Street Journal* (Oct. 6, 2023), <https://www.wsj.com/tech/ai/how-nvidia-got-hugeand-almost-invincible-da74cae1> (“CUDA gives Nvidia a competitive moat that competitors will find difficult to cross.”).

foundation model and the chip used in PC applications. If the foundation model is trained on open frameworks instead of CUDA, it will be easier to run at the edge on non-NVIDIA hardware, which may spur innovation for gen AI at the edge.

IV. Open Access to Data for AI Developers is Critical to a Robust and Open AI Ecosystem

Data is a necessary resource for the development of AI models and applications,¹⁰ and it is important that AI developers have access to data, in a manner consistent with all applicable laws. [Redacted] fully supports the protection of privacy, data security, and transparency in the development and implementation of gen AI. We believe our initiatives in developing gen AI at the edge are consistent with and promote these values. At the same time, the ecosystem needs access to data to train, fine-tune, and apply gen AI models. The possession of large amounts of data is not unlawful in itself, but restricting access to data may inhibit the development of the entire ecosystem.¹¹ Again, it is too early to tell how the companies currently collecting vast amounts of data, including through their digital ecosystems, will address the question of gen AI developers' access to data. However, given the importance of data throughout the gen AI development and deployment process, delayed or denied access to data may stunt innovation in this segment in the long term.

Accordingly, we encourage the Commission to begin investigating mechanisms to maintain open access to data for gen AI development in a manner consistent with fundamental values like privacy and transparency.

We thank the Commission for the opportunity to comment and stand ready to provide additional information or assistance as needed.

¹⁰ See Benoît Coeuré, President of the French Competition Authority, Artificial intelligence: making sure it's not a walled garden, Keynote address at the Bank for International Settlements – Financial Stability Institute policy implementation meeting on big techs in insurance (Mar. 19, 2024) https://content.mlex.com/Attachments/2024-03-19_P2ISG35N2MYY4QW4%2f20240319-BIS-Speech.pdf (“Access to massive amounts of data is key to train and fine-tune generative AI.”); FTC Gen AI Blog Post (“Developing generative AI typically requires exceptionally large datasets, especially in the pre-training step.”).

¹¹ See *id.* (“[E]ven with responsible data collection practices in place, companies' control over data may also create barriers to entry or expansion that prevent fair competition from fully flourishing.”).