



June 21, 2023

The Honorable Lina Khan
Chair, Federal Trade Commission
600 Pennsylvania Avenue, NW
Washington, DC 20580

Re: Request for Information, FTC-2023-0028

Dear Chair Khan,

We thank the Federal Trade Commission for undertaking a Solicitation for Public Comments on the Business Practices of Cloud Computing Providers (FTC-2023-0028). The Center for American Progress (CAP) is an independent, nonpartisan policy institute that is dedicated to improving the lives of all Americans through bold, progressive ideas, as well as strong leadership and concerted action.

As the landmark CAP 2021 report “How to Regulate Tech: A Technology Policy Framework for Online Services” noted about cloud infrastructure:¹

Online services provide essential infrastructure for the American economy, culture, and society. Cloud infrastructure, content delivery networks (CDNs), web hosts, and data analytics services are the quiet but dynamic backbone of the commercial internet. These online services now act as infrastructural components to other economic and social activity. They are generally lower on the stack than the more consumer-facing services at the top of the stack...They are constant and ubiquitous, even if invisible to most internet users, and for this reason are often overlooked by regulators and the public until something goes wrong. A special focus on online infrastructural services is warranted due to their tremendous impact on the economy, the environment, cybersecurity, human rights, and freedom of expression.

A better understanding of the critical Cloud Computing Providers market is warranted. CAP appreciates the request for information and respectfully offers selected focus areas and questions for your consideration.

¹ Erin Simpson and Adam Conner, “How To Regulate Tech: A Technology Policy Framework for Online Services” (Washington: Center for American Progress, 2021), available at <https://www.americanprogress.org/article/how-to-regulate-tech-a-technology-policy-framework-for-online-services/>.

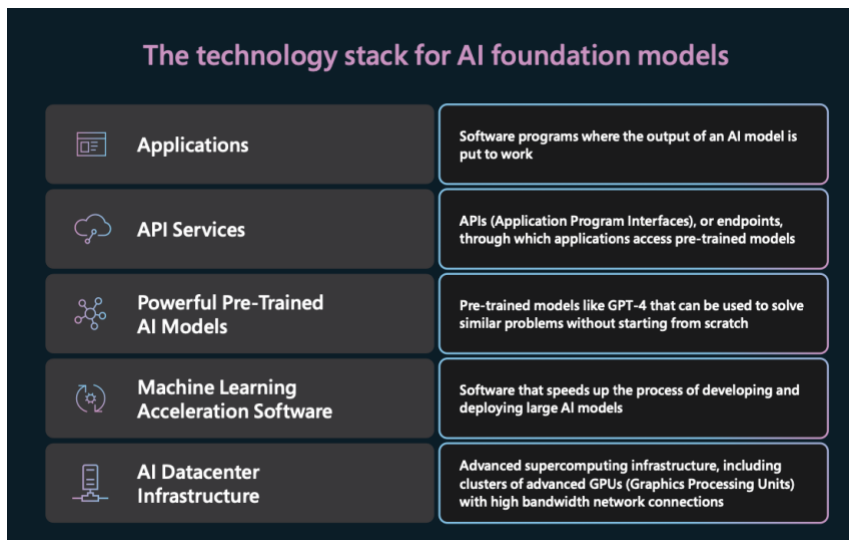
Sincerely,

Adam Conner
Vice President, Technology Policy,
Center for American Progress
aconner@americanprogress.org

Understanding the various layers of the AI cloud computing stack

The FTC should seek to better understand the ways in which the various layers of the cloud computing stack power artificial intelligence (AI), especially for advanced artificial intelligence such as foundation models or large language models (LLM)² that power recent consumer AI advancements like generative OpenAI’s ChatGPT4 AI chatbot.³

Microsoft’s May 2023 “Governing AI: A Blueprint for the Future” report includes a useful graphic that identifies “The technology stack for AI foundation models,” which includes Applications, API Services, Powerful Pre-Trained AI Models, Machine Learning Acceleration Software, and AI Datacenter Infrastructure.⁴ Each of these layers represents a distinct component required to run AI and services that are offered by major cloud computing providers.⁵



² Helen Toner, “What Are Generative AI, Large Language Models, and Foundation Models?,” *Center for Security and Emerging Technology*, May 12, 2023, available at <https://cset.georgetown.edu/article/what-are-generative-ai-large-language-models-and-foundation-models/>.

³ OpenAI, “ChatGPT,” available at <https://openai.com/chatgpt> (last accessed June 21, 2023).

⁴ Microsoft, “Governing AI: A Blueprint for the Future” (2023), available at <https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW14Gtw>

⁵ Microsoft, “Azure AI infrastructure,” available at <https://azure.microsoft.com/en-us/solutions/high-performance-computing/ai-infrastructure/#features> (last accessed June 21, 2023); Google Cloud, “AI and machine learning products,” available at <https://cloud.google.com/products/ai> (last accessed June 21, 2023); Amazon, “Machine Learning on AWS,” available at <https://aws.amazon.com/machine-learning/> (last accessed June 21, 2023).

⁶ Microsoft, “Governing AI: A Blueprint for the Future.”

Many major cloud computing providers offer commercial products at every layer of the AI cloud computing stack. This leads to important questions about the relationship between these layers. For example, does a cloud computing provider selling all five layers of the AI cloud computing stack allow its AI Datacenter Infrastructure to install and run a third-party AI model that a customer brings to their service? Or does a cloud computing provider only allow their AI Datacenter Infrastructure to run their AI model? Outlining these relationships between services offered by cloud computing providers in every layer of the AI cloud computing stack would be an important illumination.

While this example has focused on the AI cloud computing stack of cloud computing providers, similar layers exist in every part of the broader cloud computing services market. These cloud computing layers should be broadly researched, and these relationships better understood. The commercialization of each layer of the cloud computing stack is foundational to development of Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS), and their business model.

The FTC should seek to better understand the relationship between offerings of cloud computing providers at each level of the AI cloud computing stack. This understanding should include:

- What SaaS, PaaS, and IaaS services do major cloud computing providers provide at every level of the AI cloud computing stack?
- What technical interoperability or restrictions and legal/contractual terms do cloud computing providers place on each of those services at every level of the stack and the interaction of that service with other levels of AI cloud computing stack?
- What technical optimizations, financial incentives, liability restrictions, or other contractual restrictions do cloud computing providers put in place to incentivize the use of their systems at every layer of the AI cloud computing stack?
- What are the estimated increases in power usage from traditional cloud computing for each level of the AI cloud computing stack? What are the overall climate impacts from the AI computing stack? Is any increased climate impact from the AI cloud computing stack being clearly disclosed to the public? Have cloud computing providers updated their scope 1, 2, and 3 emission projections to account for AI's increase?

The importance of cloud computing providers and AI datacenter infrastructure

It takes a massive amount of computing power from specialized GPU processors to train advanced AI like foundation models or LLMs.⁷ Though it takes less computing power to run individual queries of a trained model, if the model is widely used by millions of consumers

⁷ Jared Kaplan and others, "Scaling Laws for Neural Language Models," *arXiv preprint arXiv:2001.08361*(2020): available at <https://arxiv.org/abs/2001.08361>; Andrew J. Lohn and Micah Musser, "AI and Compute: How Much Longer Can Computing Power Drive Artificial Intelligence Progress?" (Washington: Center for Security and Emerging Technology), available at <https://cset.georgetown.edu/publication/ai-and-compute/>; Amba Kak and Sarah Myers West, "2023 Landscape: Confronting Tech Power Executive Summary" (AI Now, 2023), available at <https://ainowinstitute.org/general/2023-landscape-executive-summary>.

inputting queries, the computing expenditure can be much higher in aggregate than training a model.⁸

This means using a foundation model or LLM AI requires access to significant amounts of computing power, often through specialized high-powered computing using GPU chips.⁹ There are tremendous capital costs to amass that high computing power by securing the necessary GPUs, installing and configuring them in a data center, and paying for the energy required to run them. For example, analysts estimate training a LLM such as ChatGPT-3 (not even the most advanced LLM) could cost more than \$4 million with “more advanced language models could cost over “the high-single-digit millions” to train, said Rowan Curran, a Forrester analyst who focuses on AI and machine learning.”¹⁰ That figure is likely exclusive of the costs to purchase and build a new datacenter facility and chips but is an estimate of the cost to purchase the computing power from a commercial cloud computing provider with AI datacenter infrastructure.¹¹ Additionally, the high computing power required for artificial intelligence, and the vastly increased demand for that high computing power, means a significant overall increase in the energy usage at a time when that could have a significant impact on climate change efforts.¹²

Building the capacity necessary to train and run advanced AI models, such as foundation models, comes with tremendous costs. For example, Elon Musk, one of the world’s wealthiest people,¹³ has been attempting to create new advanced AI models after conflicts with OpenAI, which he had initially funded.¹⁴ To do so he has reportedly purchased 10,000 GPUs to develop a generative AI model.¹⁵ The cost alone of purchasing those GPUs is estimated to be tens of

⁸ Jeffrey Dastin and Stephen Nellis, “For tech giants, AI like Bing and Bard poses billion-dollar search problem,” *Reuters*, February 23, 2023, available at <https://www.reuters.com/technology/tech-giants-ai-like-bing-bard-poses-billion-dollar-search-problem-2023-02-22/>.

⁹ Lohn and Musser, “AI and Compute: How Much Longer Can Computing Power Drive Artificial Intelligence Progress?”

¹⁰ Jonathan Vanian and Kif Leswing, “ChatGPT and generative AI are booming, but the costs can be extraordinary,” *CNBC*, March 13, 2023, available at <https://www.cnbc.com/2023/03/13/chatgpt-and-generative-ai-are-booming-but-at-a-very-expensive-price.html>.

¹¹ Lohn and Musser, “AI and Compute: How Much Longer Can Computing Power Drive Artificial Intelligence Progress?”

¹² Kate Saenko, “A Computer Scientist Breaks Down Generative AI’s Hefty Carbon Footprint,” *Scientific American*, May 25, 2023, available at <https://www.scientificamerican.com/article/a-computer-scientist-breaks-down-generative-ais-hefty-carbon-footprint/>.

¹³ Michelle Toh, “Elon Musk is once again the world’s richest man,” *CNN*, May 31, 2023, available at <https://www.cnn.com/2023/05/31/business/elon-musk-worlds-richest-man-intl-hnk/index.html>.

¹⁴ Lora Kolodny, “Elon Musk plans ‘TruthGPT’ A.I. to rival OpenAI, DeepMind,” *CNBC*, April 18, 2023, available at <https://www.cnbc.com/2023/04/18/musk-calls-plans-truthgpt-ai-to-rival-openai-deepmind.html#:~:text=Tesla%2C%20SpaceX%20and%20Twitter%20CEO,and%20Google%27s%20DeepMind%2C%20he%20said.>

¹⁵ Kali Hays, “Elon Musk is moving forward with a new generative-AI project at Twitter after purchasing thousands of GPUs,” *Business Insider*, April 11, 2023, available at <https://www.businessinsider.com/elon-musk-twitter-investment-generative-ai-project-2023-4>.

millions of dollars.¹⁶ Musk already has access to existing data centers to install these GPUs thanks to his purchase of Twitter; the company would also be on the hook for the potentially considerable power bill for running that many GPUs.

The high capital costs of this intensive computing power plus the limited availability of GPUs means that most of a company's advanced AI usage will be done on AI datacenter infrastructure or advanced AI models from a handful of cloud computing providers using their commercially available IaaS or PaaS offerings for deep learning or AI.¹⁷

As the 2023 AINow Landscape report described this phenomenon, "Only a handful of companies actually run their own infrastructure – the cloud and computer resources foundational to building AI systems. What this means is that even though 'AI startups' abound, they must be understood as barnacles on the hull of Big Tech – licensing server infrastructure."¹⁸

Another analogy may be the mobile operating system ecosystem. While there are hundreds of thousands of mobile applications available there are only two mobile operating systems, Android and Apple iOS, which control the distribution of these apps and the phones they run on (though the mobile apps themselves are often hosted on cloud computing providers). There may be numerous AI startups and AI models, but most of them are going to be running on the AI datacenter infrastructure of a few cloud computing providers.

At the same time, there are reports that major cloud computing providers are struggling to meet the internal demand from their own employees to access AI computing power to develop or run their own AI products.¹⁹ If major cloud computing providers are struggling to meet their own demand for AI computing power from AI datacenter infrastructure, that can mean a shortage or inability to provide AI datacenter infrastructure to existing or new customers.²⁰ The inability of a customer to access sufficient computing power on AI datacenter infrastructure from a major cloud computing provider means that company may not be able to develop new AI models or run existing or new models in production.

¹⁶ Benj Edwards, "Elon Musk reportedly purchases thousands of GPUs for generative AI project at Twitter," *Ars Technica*, April 12, 2023, available at <https://arstechnica.com/information-technology/2023/04/elon-musk-reportedly-purchases-thousands-of-gpus-for-generative-ai-project-at-twitter/#:~:text=Despite%20recently%20calling%20for%20a,people%20familiar%20with%20the%20company.>

¹⁷ Run.AI, "Cloud Deep Learning: Top Three Platforms Compared," available at <https://www.run.ai/guides/cloud-deep-learning>.

¹⁸ Amba Kak and Sarah Myers West, "2023 Landscape: Confronting Tech Power" (AINow, 2023), available at <https://ainowinstitute.org/wp-content/uploads/2023/04/AI-Now-2023-Landscape-Report-FINAL.pdf> <https://ainowinstitute.org/wp-content/uploads/2023/04/AI-Now-2023-Landscape-Report-FINAL.pdf>.

¹⁹ Aaron Holmes and Kevin McLaughlin, "Microsoft Rations Access to AI Hardware for Internal Teams," *The Information*, March 15, 2023, available at <https://www.theinformation.com/articles/microsoft-rations-access-to-ai-hardware-for-internal-teams?rc=cwjrii>.

²⁰ Aaron Holmes and Anissa Gardizy, "AI Developers Stymied by Server Shortage at AWS, Microsoft, Google," *The Information*, April 7, 2023, available at <https://www.theinformation.com/articles/ai-developers-stymied-by-server-shortage-at-aws-microsoft-google?rc=cwjrii>.

Some new cloud computing providers have arisen to meet this market demand.²¹ But there is a limited supply of these specialized GPUs due to the industry's insufficient capacity to meet recent demand (the high computing needs of cryptocurrency and AI, supply chain constraints, and limited manufacturing capacity.)²² Existing large cloud computing providers are using their cash advantages to sign deals with startups that have additional capacity in their AI datacenter infrastructure,²³ meaning it could be even harder for small to medium companies to access the necessary AI computing power.

The FTC should seek to better understand the impact of major cloud computing providers and their AI datacenter infrastructure being the primary, or only, option for most entities to develop and run advanced AI models including:

- What are the AI datacenter infrastructure options available to a new startup to train or run a new foundation AI model?
- What would the estimated capital costs be in establishing the computing power needed to train or run a new AI foundation model?
- Given heavy demand and supply chain limitations, is it even possible for a new entrant with sufficient capital to purchase enough GPUs to establish the capacity to train or run a new foundation AI model? What is the expected timeline for that to change?
- What are the main AI datacenter infrastructure commercial offerings (SaaS, PaaS, and IaaS) from major cloud computing providers?
- Do major cloud computing providers allow their commercially available AI datacenter infrastructure to be used to train AI models? To run AI models that were not developed approved by the cloud Computing Provider?
- Do major cloud computing providers share the same AI datacenter infrastructure for their own first-party AI model usage as they sell for commercial third-party usage? If so, how is computing power allocated between the provider's first-party use and commercial third-party demand?
- How do changes to AI datacenter infrastructure and increased AI demand effect the climate impact of data center for cloud computing providers?
- How do the increased power requirements for AI datacenter infrastructure impact the power grid? How might this impact the reliability and safety of the power grid? Does increased power usage create a critical infrastructure concern for either the power grid or the cloud computing services that run through major cloud computing providers? Are there localized impacts that to regional power grids that could impose risks to household and business consumers?

²¹ Holmes and Gardizy, "AI Developers Stymied by Server Shortage at AWS, Microsoft, Google."

²² Reuters, "Explainer: Why are Nvidia's shares soaring and what is its role in the AI boom," *Reuters*, June 14, 2023, available at <https://www.reuters.com/technology/why-are-nvidias-shares-soaring-what-is-its-role-ai-boom-2023-05-25/#:~:text=May%2030%20%28Reuters%29%20-%20Nvidia%20Corp%20%28NVDA.O%29%20soared,for%20the%20chipmaker%20from%20an%20artificial%20intelligence%20boom.>

²³ Jordan Novet, "Microsoft signs deal for A.I. computing power with Nvidia-backed CoreWeave that could be worth billions," *CNBC*, June 1, 2023, available at <https://www.cnbc.com/2023/06/01/microsoft-inks-deal-with-coreweave-to-meet-openai-cloud-demand.html>.

U.S. Government Security Requirements and cloud computing providers

The adoption of cloud computing has only accelerated as private sector entities became comfortable with the security features of these commercial SaaS, PaaS, and IaaS offerings. Question 17 asks “What effect does security-related regulation (e.g., GLBA Safeguards Rule, HIPAA Security Rule, etc.) have on market dynamics?” Part of the concentration in cloud computing providers can be traced to the cost of the investment for additional compliance measures required by laws like the Health Insurance Portability and Accountability Act (HIPAA) or Payment Card Industry (PCI) requirements.²⁴

However, it should not be noted that in the absence of needed new cybersecurity laws or regulations, that some of the higher security offerings and certifications for commercial cloud computing providers come directly from the federal government’s own requirements for government cloud computing security use, which has a significant effect on the market dynamics, as well.²⁵

The General Service Administration’s Federal Risk and Authorization Management Program (FedRAMP)²⁶ established in 2011 created a process to allow for the authorization of commercial cloud services for use by federal agencies and the public listing of products with those authorizations.²⁷ This includes the authorization for moderate and high-impact levels for the most common commercial offerings from major cloud computing providers in SaaS, PaaS, and IaaS²⁸ and specialized government only cloud computing products.²⁹ Similarly, major cloud computing providers have Cloud Computing Authorizations³⁰ from the Department of Defense through the DoD Cloud Authorization Services (DCAS) at the Defense Information Systems Agency (DISA).³¹ Additionally, the selection of four major cloud computing providers for the

²⁴ Amazon, “AWS Cloud Security: HIPAA,” available at <https://aws.amazon.com/compliance/hipaa-compliance/>; Microsoft, “PCI DSS,” available at <https://learn.microsoft.com/en-us/azure/compliance/offerings/offering-pci-dss>.

²⁵ Federal Cloud Computing Strategy, “From Cloud First to Cloud Smart,” available at <https://cloud.cio.gov/strategy/>.

²⁶ FedRAMP, “Home,” available at <https://www.fedramp.gov>.

²⁷ “FedRAMP Marketplace,” *FedRAMP*, available at <https://marketplace.fedramp.gov/products>.

²⁸ “AWS US East/West,” available at <https://marketplace.fedramp.gov/products/AGENCYAMAZONEW>; “Azure Commercial Cloud,” available at <https://marketplace.fedramp.gov/products/F1209051525>; “Google Workspace,” available at <https://marketplace.fedramp.gov/products/F1206081364>;

“Google Services (Google Cloud Platform Products and underlying Infrastructure),” available at <https://marketplace.fedramp.gov/products/FR1805751477>; “Office 365 Multi-Tenant & Supporting Services,” available at <https://marketplace.fedramp.gov/products/MSO365MT>.

²⁹ “AWS GovCloud,” available at <https://marketplace.fedramp.gov/products/F1603047866>;

“Azure Government (includes Dynamics 365),” available at <https://marketplace.fedramp.gov/products/F1603087869>;

“Microsoft Office 365 GCC High,” available at <https://marketplace.fedramp.gov/products/FR1824057433>.

³⁰ Microsoft, “Department of Defense (DoD) Impact Level 5 (IL5),” available at <https://learn.microsoft.com/en-us/compliance/regulatory/offering-dod-il5>.

³¹ U.S. Department of Defense Cyber Exchange, “DoD Cloud Authorization Process,” available at <https://public.cyber.mil/dccs/dod-cloud-authorization-process/>.

Department of Defense's Joint Warfighting Cloud Capability Procurement contracts,³² after an extended contracting process, places those cloud computing providers in a unique category.

These federal government certifications act as a signal of good security practices to both public and private sector actors looking to purchase commercial cloud services. It serves as a definitive stamp of security for a corporate CIO or CISO that is otherwise difficult to obtain in the private sector through contracts or audits alone. Now, to be clear, not every commercial service from a cloud computing provider provides the same level of security as the specific service that is certified by a program like FEDRAMP (though some like Amazon have certified their entire AWS US East/West service for moderate impact level).³³ But the ability to comply with and certify U.S. federal government cloud security programs is a demonstration of clear compliance, legal, and security capacity.

While the government cloud security authorization process is less strenuous than it was previously and a new law pushes agencies to accept existing authorizations,³⁴ there are still high initial certification and ongoing compliance costs with these requirements. This means, for example, in response to Question 11 in the RFI that the federal government is reliant on "a small handful of cloud service providers," it is because only a few cloud computing providers who have chosen to undergo the process required to sell to the federal government.

The FTC should seek to better understand the role that federal government cloud security requirements play on impacting the market for cloud computing providers and what security requirements could or should provide broader assurance or access for commercial customers including:

- What impact do the U.S. government cloud security requirements have on the cloud computing market?
- Should certifications for cloud security programs like FEDRAMP be required to denote what protections are provided for all commercial customers versus those solely for government customers to provide the ability for clearer standardized comparison?
- To what extent do multi-provider cloud computing contracts like the DOD's Joint Warfighting Cloud Capability Procurement encourage or require interoperability between multi-cloud environments from major commercial cloud providers? If so, could any multi-cloud interoperability between major commercial cloud providers selected for the contract be encouraged or provided to the broader commercial customer based?
- Do vastly increased AI computing requirements on AI datacenter infrastructure (described above) cause potential risk to the ability for cloud computing providers to

³² U.S. Department of Defense, "Department of Defense Announces Joint Warfighting Cloud Capability Procurement," Press release, December 7, 2022, available at <https://www.defense.gov/News/Releases/Release/Article/3239378/department-of-defense-announces-joint-warfighting-cloud-capability-procurement/>

³³ "AWS US East/West," available at <https://marketplace.fedramp.gov/products/AGENCYAMAZONEW>.

³⁴ James M. Inhofe National Defense Authorization Act for Fiscal Year 2023, H.R. 7776, 117th Cong., 2nd sess. (January 3, 2022), available at <https://www.congress.gov/117/bills/hr7776/BILLS-117hr7776enr.pdf#page=1055>.

provide contracted cloud computing services to the U.S. government, if those facilities or services are shared or co-located?