

# Comments on the Business Practices of Cloud Computing Providers

Filed June 21, 2023 before the Federal Trade Commission, Docket ID FTC-2023-0028

Jai Vipra, Research Fellow and Sarah Myers West, Managing Director, AI Now Institute

## Table of Contents

1. Executive Summary
2. The highly concentrated cloud computing market is contributing to concentration in the AI industry.
3. Effective policy interventions must attend to concentration across layers of the computing stack.
4. Cloud providers are best positioned to reap the advantages of collecting vast amounts of data, which can be used to reinforce their dominant position in AI.
5. Suggested points for further consideration

## Executive Summary

In its request for information, the Federal Trade Commission has invited public comments on the Cloud Computing Industry and the practices of cloud computing providers. One area of particular interest to the Federal Trade Commission relates to market power and business practices affecting competition in cloud computing, and to what extent the use of AI affects competition among cloud providers today and into the future.

This response outlines that the effects of concentration in the cloud computing market and the development of artificial intelligence are profound and mutually reinforcing. The overwhelming trajectory of AI development over the past decade has trended toward larger and larger scale, creating dependencies on massive amounts of data and the computational

power to process this data.<sup>1</sup> These dependencies give significant market power to the small handful of companies which have amassed control over these resources - cloud infrastructure firms that also operate widespread platform ecosystems, and which have first mover advantage in large-scale artificial intelligence.<sup>2</sup> This reinforces the findings of the House Digital Markets Report, which underscored that market leaders in cloud computing benefit from early-mover advantages due to network effects and high switching costs.<sup>3</sup>

This holds true even as new start-ups begin to build commercial AI products: to enter the field, small companies must secure compute credits or make other contractual arrangements with Big Tech firms.<sup>4</sup> These companies face significant barriers to entry if they were to seek to build these resources from scratch, due significant start-up costs, lack of interoperability at key points in the compute stack and bottlenecks in the supply chain for key components of compute infrastructure. Talent requirements also grow as compute costs grow, because very specialized knowledge is needed to make the most of scarce hardware, and much of this knowledge is tacit. Talent is thus a significant barrier to entry related to high compute costs.<sup>5</sup>

These effects work in combination to more deeply entrench the infrastructural and economic power of the few firms that retain control over the key components to building AI, with detrimental effects on competition in the AI industry. This also contributes to consumer injury in many forms, including harms to privacy and security,<sup>6</sup> encouraging the spread of false and misleading information,<sup>7</sup> perpetuating patterns of inequality and discrimination,<sup>8</sup> harmful effects on workers,<sup>9</sup> and environmental harms.<sup>10</sup>

---

<sup>1</sup> Amba Kak and Sarah Myers West, "Confronting Tech Power", AI Now Institute, April 2023, <https://ainowinstitute.org/2023-landscape>.

<sup>2</sup> Sarah Myers West, 'Competition Authorities Need to Move Fast and Break up AI', Financial Times, 17 April 2023, <https://www.ft.com/content/638b5be7-fab7-4fe6-a0cf-7dabefcdd722>

<sup>3</sup> Investigation of Competition in Digital Markets, Subcommittee on Antitrust, Commercial and Administrative Law of the Committee on the Judiciary of the House of Representatives, 117th Congress, July 2022, p. 94,

<https://www.govinfo.gov/content/pkg/CPRT-117HPRT47832/pdf/CPRT-117HPRT47832.pdf/>

<sup>4</sup> See, for example, OpenAI's arrangement with Microsoft, HuggingFace and Stable Diffusion's arrangements with Amazon Web Services, and Anthropic's arrangement with Google Cloud.

<sup>5</sup> Micah Musser et al., 'The Main Resource Is the Human', Center for Security and Emerging Technology, April 2023, <https://cset.georgetown.edu/publication/the-main-resource-is-the-human/>

<sup>6</sup> Epic.org, New EPIC Report Sheds Light on Generative A.I. Harms, <https://epic.org/new-epic-report-sheds-light-on-generative-a-i-harms/>

<sup>7</sup> Michael Atleson, 'The Luring Test: AI and the engineering of consumer trust', Federal Trade Commission, 1 May 2023, <https://www.ftc.gov/business-guidance/blog/2023/05/luring-test-ai-engineering-consumer-trust>

<sup>8</sup> Elisa Jillson, 'Aiming for truth, fairness, and equity in your company's use of AI', Federal Trade Commission, 19 April 2021, <https://www.ftc.gov/business-guidance/blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai>

<sup>9</sup> Wilneida Negrón, 'Little Tech is Coming for Workers', Coworker.org, November 2021, <https://home.coworker.org/wp-content/uploads/2021/11/Little-Tech-Is-Coming-for-Workers.pdf>

<sup>10</sup> Emily Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?.' In Proceedings of the 2021 ACM conference on

Drawing on forthcoming work, this response traces through the effects of concentration in cloud computing on artificial intelligence across three levels: the infrastructural dominance of cloud infrastructure providers, competition concerns that surface across layers of the technical stack,<sup>11</sup> and harms related to the significant data advantages of these companies.<sup>12</sup> We conclude with a list of potential avenues for further consideration by the Federal Trade Commission and points of intervention:

### **Suggested points for further consideration:**

1. To address the most pressing concerns about current and future anti-competitive behavior in the AI market, the Federal Trade Commission should explore structural interventions targeted to the largest cloud infrastructure companies and their commercial AI model offerings.
2. The Federal Trade Commission should also enforce relevant antitrust laws against likely anti-competitive behavior, including restrictive licensing regimes and tying/bundling practices at different layers of the tech stack.
3. The FTC should clarify the legality of data use for model training, including whether the use of data by cloud providers to train their AI models violates consumer control over personal data and whether such data use should be considered in the FTC's commercial surveillance Advanced Notice of Proposed Rulemaking.

---

fairness, accountability, and transparency, pp. 610-623. 2021.

<https://dl.acm.org/doi/10.1145/3442188.3445922>

<sup>11</sup> For example, in addition to contractual agreements for compute power, a limited number of firms are able to control certain layers of the compute stack through bundling software and hardware, as outlined below.

<sup>12</sup> Lina M. Khan, 'Amazon's Antitrust Paradox', 1 January 2017,

[https://www.yalelawjournal.org/pdf/e.710.Khan.805\\_zuvfyeh.pdf](https://www.yalelawjournal.org/pdf/e.710.Khan.805_zuvfyeh.pdf); Seeta Peña Gangadharan, Working Around Democracy: Big Tech, Computational Power, and Racial Equity, PolicyLink,

; <https://bayareaequityatlas.org/sites/default/files/DrSeeta.pdf>; Seda Gurses and Joris van Hoboken, 'Privacy after the Agile Turn', SocArXiv, 2 May 2017, <https://osf.io/preprints/socarxiv/9gy73/>

## A highly concentrated cloud computing market is contributing to concentration in the AI industry.

Training and operating large-scale AI systems is reliant on computational infrastructure (“compute”). This means that concentration in the cloud computation market influences concentration in AI broadly. Compute is a key bottleneck in both the creation (training) of AI models and the deployment of those models in AI products (often referred to in the literature as ‘inference’, or when the model is asked to infer things about new data based on its training). This dependency on access to compute gives three cloud infrastructure providers, Amazon, Google, and Microsoft, an advantage in AI development and deployment.

The focus on compute power in AI is a product of conclusions drawn in the AI research literature that AI model capabilities improve as scale increases.<sup>13</sup> This has driven exponentially growing demand for computing power, leading to concerns that the current pace of growth is unsustainable.<sup>14</sup> At these large scales, compute costs are, in the words of Sam Altman, “eye watering”. The final training run of GPT-4 is estimated to have cost \$40 million, though the overall training cost may in fact be much higher because compute is also required for trial and error before the final training run.<sup>15</sup>

While Alphabet, OpenAI and other tech companies have signaled a need to increase training efficiency to save on compute costs,<sup>16</sup> the behavior exhibited by these companies suggests otherwise. Rather than seek improvements to the underlying efficiency of their models, cloud infrastructure providers have instead implemented temporary workarounds for the short-term and made investments in proprietary chips that would lock in their infrastructural dominance in the long term.

Both OpenAI and DeepMind sought (different levels of) integration with the technology giants Microsoft and Alphabet due to their inability to provide the requisite compute for training large

---

<sup>13</sup> Neil C. Thompson, Kristjan Greenewald, Keeheon Lee, and Gabriel F. Manso, ‘The Computational Limits of Deep Learning’, arXiv, 27 July 2022, <http://arxiv.org/abs/2007.05558>; Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei, ‘Scaling Laws for Neural Language Models’, arXiv, 23 January 2020, <https://arxiv.org/abs/2001.08361>

<sup>14</sup> Andrew J. Lohn and Micah Musser, ‘AI and Compute: How Much Longer Can Computing Power Drive Artificial Intelligence Progress?’, Center for Security and Emerging Technology, January 2022, [https://cset.georgetown.edu/wp-content/uploads/AI-and-Compute-How-Much-Longer-Can-Computing-Power-Drive-Artificial-Intelligence-Progress\\_v2.pdf](https://cset.georgetown.edu/wp-content/uploads/AI-and-Compute-How-Much-Longer-Can-Computing-Power-Drive-Artificial-Intelligence-Progress_v2.pdf)

<sup>15</sup> Ben Cottier, ‘Trends in the dollar training cost of machine learning systems’, Epoch, 31 January 2023, <https://epochai.org/blog/trends-in-the-dollar-training-cost-of-machine-learning-systems>

<sup>16</sup> Hayden Field, ‘Tech’s new business model: ‘Do more with less’’, CNBC, 11 May 2023, <https://www.cnbc.com/2023/05/11/techs-new-business-model-do-more-with-less-.html>

models.<sup>17</sup> Since the launch of Office 365 tools powered by GPT-4, Microsoft has reportedly rationed access to its AI hardware for internal teams<sup>18</sup>, while Alphabet combined its Google Brain and DeepMind teams to ensure greater resource sharing internally.<sup>19</sup> While the market for the most advanced AI chips is experiencing a supply shortage,<sup>20</sup> cloud providers have an incentive to reserve scarce compute for their own models, or for the models of companies they have investments in, which may have downstream negative effects on cloud customers.

The overreliance on AI at scale is leading to toxic competition<sup>21</sup> exhibited by the rush to deploy systems without adequate underlying infrastructure, guardrails, or pre-testing in order to retain first-mover advantage,<sup>22</sup> experimenting in the wild rather than taking steps to ensure consumer protections are in place and that systems are compliant with existing law. AI at scale is also associated with ancillary harms, including exacerbation of labor harms, environmental damage, discriminatory outputs, and the proliferation of hate speech.<sup>23</sup>

**Demand for compute is shaping the trajectory of academic AI research, and reinforcing industry-led approaches. Open source software does not, at present, offer meaningful alternatives.**

Over the past decade the balance of research in the field of artificial intelligence has tipped heavily toward increasing industry dominance, driven by demand for computing power and

---

<sup>17</sup> 'OpenAI LP', OpenAI Blog, <https://openai.com/blog/openai-lp>; Jeremy Kahn, 'The Google Brain-DeepMind merger is good for Google. It might not be for us' Fortune, 29 April 2023, <https://fortune.com/2023/04/28/the-google-brain-deepmind-merger-alphabet-pichai-risks-eye-on-ai/>

<sup>18</sup> Aaron Holmes and Kevin McLaughlin, 'Microsoft Rations Access to AI Hardware for Internal Teams', The Information, 15 March 2023, <https://www.theinformation.com/articles/microsoft-rations-access-to-ai-hardware-for-internal-teams?rc=7gpwfr>

<sup>19</sup> 'Announcing Google DeepMind', 20 April 2023, <https://www.deepmind.com/blog/announcing-google-deepmind>

<sup>20</sup> Deepa Seetharaman and Tom Dotan, 'The AI Boom Runs on Chips, but It Can't Get Enough', The Wall Street Journal, 29 May 2023, <https://www.wsj.com/articles/the-ai-boom-runs-on-chips-but-it-cant-get-enough-9f76f554>; Dan Gallagher, 'AI Spending Will Cloud Chip Slowdown', The Wall Street Journal, 15 May 2023, <https://www.wsj.com/articles/ai-spending-will-cloud-chip-slowdown-1dbf261d>

<sup>21</sup> Ariel Ezrachi and Maurice E. Stucke, 'The Curious Case of Competition and Quality', Journal of Antitrust Enforcement, Volume 3, Issue 2, pp. 227-257, October 2015, <https://doi.org/10.1093/jaenfo/jnv023>

<sup>22</sup> Nico Grant and Cade Metz, 'A New Chat Bot Is a 'Code Red' for Google's Search Business', The New York Times, 21 December 2022, <https://www.nytimes.com/2022/12/21/technology/ai-chatgpt-google-search.html>

<sup>23</sup> AI Now Institute, 'ChatGPT And More: Large Scale AI Models Entrench Big Tech Power', 2023 Landscape, 11 April 2023, <https://ainowinstitute.org/publication/large-scale-ai-models>; Emily Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?.' In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency, pp. 610-623, 2021. <https://dl.acm.org/doi/10.1145/3442188.3445922>

access to data.<sup>24</sup> This has produced a self-reinforcing process, in which the dominant trends in AI research are shaped by dependency on corporate resources, which in turn reinforce this dependency by producing an academic monoculture.<sup>25</sup>

Open-source pre-trained models offer few meaningful alternatives: the largest purely academia-developed self supervised model used only 1 percent of the compute used to train the largest industry model, far below the scale of commercial-grade AI.<sup>26</sup> With few exceptions, the bulk of open source development is taking place at the level of fine-tuning models that were pre-trained using Big Tech resources, with continued need for cloud infrastructure providers' resources for inference.<sup>27</sup> Moreover, Big Tech firms have stated outright that they see the development of open source software as a path to making money by extracting rents across their platform ecosystems.<sup>28</sup>

## Effective policy interventions must attend to concentration across layers of the computing stack.

In addition to concerns about market dominance in AI by cloud providers, regulators should be concerned about concentration within different layers of the cloud computing stack.

Compute is a stack that includes hardware, software and infrastructure. In addition to the amount of compute that any individual player can leverage, being behind the curve on any component in this stack means that AI developers will lag behind their competitors. Many

---

<sup>24</sup> Brian Eastwood, 'Study: Industry now dominates AI research', Ideas Made to Matter, MIT Sloan Management School, 18 May 2023,

<https://mitsloan.mit.edu/ideas-made-to-matter/study-industry-now-dominates-ai-research>

<sup>25</sup> Meredith Whittaker, 'The steep cost of capture', Interactions, Volume 28, Issue 6, pp. 50-55, 10 November 2021, <https://dl.acm.org/doi/10.1145/3488666>; Mohamed Abdalla and Moustafa Abdalla, The Grey Hoodie Project: Big Tobacco, Big Tech, and the threat on academic integrity, arXiv, 27 April 2021, <https://arxiv.org/abs/2009.13676>

<sup>26</sup> Tamay Besiroglu, Sage Andrus Bergerson, Amelia Michael, Xueyun Luo, Lennart Heim, and Neil Thompson. The Compute Divide in Machine Learning: A Threat to Academic Contribution and Scrutiny? Forthcoming.

<sup>27</sup> Exceptions to this include Eleuther, which builds its models through an arrangement with Coreweave, HuggingFace's BLOOM model, which was trained on a supercomputer funded by the French government (but has a revenue-sharing agreement with AWS for ongoing use of its resources). The current best-in-class open source model, Falcon 50-B, was trained using AWS. The dependencies on Big Tech resources also extend to data (for example, the CommonCrawl dataset used in the training of many large AI models is hosted by Amazon Web Services' Open Data Management program <https://commoncrawl.org/the-data/get-started/>) and development environments (TensorFlow and PyTorch, the two leading libraries used for machine learning development, are created and maintained by Google and Meta, respectively). More in Widder, Whittaker and West, forthcoming.

<sup>28</sup> Sylvia Varnham O'Regan, Jon Victor and Amir Efrati, 'Meta Wants Companies to Make Money Off Its Open-Source AI, in Challenge to Google', The Information, 15 June 2023, <https://www.theinformation.com/articles/meta-wants-companies-to-make-money-off-its-open-source-ai-in-challenge-to-google>; Dylan Patel and Afzal Ahmad, 'Google "We Have No Moat, And Neither Does OpenAI"', SemiAnalysis, 4 May 2023, <https://www.semianalysis.com/p/google-we-have-no-moat-and-neither>

components of this stack exhibit a high degree of market concentration, suggesting the need to attend to market dynamics both at the level of the cloud computing market, and within particular layers of the compute stack.<sup>29</sup>

The leading AI companies have all also made astronomical investments in AI chips that would further cement their position in the market.<sup>30</sup> Google, Amazon, and Microsoft are all working to bring aspects of their chip-making in-house, customizing chips specifically to their applications.<sup>31</sup> Such moves would grant each company greater control over the integration of their hardware and software, in the process creating potential ties between software and hardware elements of the compute tech stack (outlined below).

**Access to leading infrastructure is a key competitive advantage in AI. AI developers' fortunes depend on whether they are able to carve out early access deals with one of a few cloud providers (e.g. Microsoft, Amazon, or Google), who in turn must strike early access deals with one of a few chip designers or develop their own.**

### Hardware:

The hardware supply chain is remarkably concentrated, especially in the semiconductor industry including within the domains of chip design and fabrication.<sup>32</sup> Tacit knowledge, high fixed costs, and long development times are among the factors that lead to this concentration.

The competitive value of using the most capable chips cannot be overstated: state-of-the-art AI chips are usually 33 times more cost-effective than trailing AI chips.<sup>33</sup> These cost savings are driven by energy savings. Only three significant entities currently design competitive GPUs: NVIDIA, Advanced Micro Devices (AMD), and Intel.

---

<sup>29</sup> Saif M. Khan, 'The Semiconductor Supply Chain', Center for Security and Emerging Technology, January 2021, <https://cset.georgetown.edu/publication/the-semiconductor-supply-chain/>

<sup>30</sup> 'Microsoft developing its own AI chip - The Information', Reuters, 18 April 2023, <https://www.reuters.com/technology/microsoft-developing-its-own-ai-chip-information-2023-04-18/>

<sup>31</sup> Sam Shead, 'Tech giants are rushing to develop their own chips – here's why', 7 September 2021, <https://www.cnbc.com/2021/09/06/why-tesla-apple-google-and-facebook-are-designing-their-own-chips.html>; Anissa Gardizy and Wayne Ma, 'Microsoft Readies AI Chip as Machine Learning Costs Surge', The Information, 18 April 2023, <https://www.theinformation.com/articles/microsoft-readies-ai-chip-as-machine-learning-costs-surge?rc=7gpwfr>

<sup>32</sup> Saif M. Khan, 'The Semiconductor Supply Chain', Center for Security and Emerging Technology, January 2021, <https://cset.georgetown.edu/publication/the-semiconductor-supply-chain/>

<sup>33</sup> Saif M. Khan, 'AI Chips: What They Are and Why They Matter', Center for Security and Emerging Technology, April 2020, <https://cset.georgetown.edu/publication/ai-chips-what-they-are-and-why-they-matter/>

Chip fabrication facilities (to physically produce GPUs) can cost tens of billions of dollars to build. There are only three chipmakers for leading node chips: Taiwan Semiconductor Manufacturing Company (TSMC), Samsung, and Intel. Of these three, it is likely only TSMC can competitively produce next-generation node chips currently.

Cloud infrastructure providers have reinvested profits into chipmaking. Of these, Google is the only one who has chips in deployment, and these are designed to be tied to Google's own frameworks for AI development and production. Microsoft has its own chip in development, Athena, but this is not projected to be in use until 2024 at the earliest.<sup>34</sup> Both companies currently also purchase chips from Nvidia, but experience shortages,<sup>35</sup> while Nvidia may be giving two smaller companies, Coreweave and Lambda Labs, preferential access to its new H100 GPUs.<sup>36</sup>

### **Data Centers:**

Chips form only one component of compute infrastructure: data centers, or the physical location used to house computing hardware and the cabling used to connect this hardware, are also critical elements of compute infrastructure. These ensure that information can be sent to and from the compute hardware with minimal impact on latency, or the speed at which data is sent to and from the cloud.

The connective cabling needed to send information to and from data centers is scarce and expensive. Copper in particular is undergoing significant shortages, driving the need to innovate using new materials.<sup>37</sup> But the largest firms are able to make these investments, enabling them to lower their land costs by placing their data centers in rural areas far away from metropolitan centers. By contrast, smaller cloud providers must locate their data centers in more geographically proximate locations to their end users, driving up land costs.

### **Software:**

Compute software is a key component of computational infrastructure and acts as a major barrier to entry for chip designers. The dominant player in this layer of the stack is Nvidia and

---

<sup>34</sup> Anissa Gardizy and Wayne Ma, 'Microsoft Readies AI Chip as Machine Learning Costs Surge', The Information, 18 April 2023,

<https://www.theinformation.com/articles/microsoft-readies-ai-chip-as-machine-learning-costs-surge?rc=7gpwfr>

<sup>35</sup> Deepa Seetharaman and Tom Dotan, 'The AI Boom Runs on Chips, but It Can't Get Enough', The Wall Street Journal, 29 May 2023,

<https://www.wsj.com/articles/the-ai-boom-runs-on-chips-but-it-cant-get-enough-9f76f554>

<sup>36</sup> Anissa Gardizy, 'Why Nvidia Aids Cloud Rivals of AWS, Google and Microsoft', The Information, 1 June 2023,

<https://www.theinformation.com/articles/why-nvidia-aids-cloud-rivals-of-aws-google-and-microsoft>

<sup>37</sup> Dylan Patel and Afzal Ahmad, 'Thick Cables, Thin Margins – Microsoft, Amazon, and Google Demand Overstated By \$CRDO Credo', SemiAnalysis, 15 February 2023,

<https://www.semianalysis.com/p/thick-cables-thin-margins-microsoft>; Yusuf Khan, 'Copper Shortage Threatens Green Transition', The Wall Street Journal, 18 April 2023,

<https://www.wsj.com/articles/copper-shortage-threatens-green-transition-620df1e5>



its CUDA software used to compile and deploy software on Nvidia's GPU hardware.<sup>38</sup> CUDA is a challenging language and increases switching and talent costs for AI development, as developers differentiate themselves by optimizing the use of expensive hardware through acumen in complex software ecosystems.

Key components of hardware and software in compute are bundled together:

- NVIDIA has created CUDA, a programming stack that can be used to optimize AI software for its chips (GPUs).
- Google has created TensorFlow, a framework that works best with Google's own AI chips (TPUs).
- Some open source alternatives are in nascent form, including Triton, an open source alternative developed by OpenAI in 2021. In theory Triton could run on any hardware system but it currently only can be run on NVIDIA's systems.<sup>39</sup>

Software forms a key component of compute: this includes the compiling software that converts source code into executable form that can run on chips, electronic design automation software used for chip design, domain-specific languages and frameworks for machine learning, and data management software. Software that works with a broader range of hardware could make competition easier in the compute market, by lowering switching costs, while proprietary and narrowly applicable software intensifies lock-in. Open source frameworks can technically be used in place of some proprietary software frameworks, but they would be impractically slow and unwieldy.

Chip designers, like Nvidia, have an incentive to build a walled garden of software around their chips so as to lock AI companies in; cloud providers have an incentive to design their own chips to reduce their own dependence on NVIDIA. AI companies that are not owned by cloud providers have an incentive to not be locked into a chip, software or cloud ecosystem, but have few alternatives at present.

---

<sup>38</sup> Dylan Patel, 'How Nvidia's CUDA Monopoly In Machine Learning Is Breaking - OpenAI Triton And PyTorch 2.0', SemiAnalysis, 16 January 2023,

<https://www.semianalysis.com/p/nvidiaopenaitritonpytorch>

<sup>39</sup> 'Introducing Triton: Open-source GPU programming for neural networks', OpenAI, 28 July 2021,

<https://openai.com/research/triton>

Cloud providers are best positioned to reap the advantages of collecting vast amounts of data, which can be used to reinforce their dominant position in AI.

Data quantity and quality is an important factor in increasing the capabilities of large AI models.<sup>40</sup> Data quality in particular may also offer a path toward increasing the efficiency of models and reducing demand on compute resources. Cloud infrastructure providers have significant access to proprietary data, making them particularly well placed to have a data advantage in the AI market. They have closely guarded information about the training data used to build their large-scale AI models, a further indication of the sensitivity of this information.

For instance, YouTube offers a rich repository of training data that Google is well-placed to leverage; OpenAI also reportedly secretly used YouTube data to train some of its models.<sup>41</sup> Google could furthermore leverage data from Google Drive and Gmail to train its AI models, especially since these models are already integrated into these services. This is important because AI companies are running out of web data that is commonly available to use in models.<sup>42</sup>

It is likely that cloud providers will determine that they are GDPR compliant if they use user data to train AI models, arguing that the data was used to improve a product or service, such that additional consent is not required. However, the FTC's recent case against Alexa provides an important precedent in its determination that Amazon could not justify retention of ill-gotten data for the purpose of product improvement.<sup>43</sup>

---

<sup>40</sup> Rishi Bommasani et. al., 'On the Opportunities and Risks of Foundation Models', arXiv, 16 August 2021, <https://arxiv.org/abs/2108.07258>

<sup>41</sup> Jon Victor, 'Why YouTube Could Give Google an Edge in AI', The Information, 14 June 2023, <https://www.theinformation.com/articles/why-youtube-could-give-google-an-edge-in-ai?rc=7gpwfr>

<sup>42</sup> Pablo Villalobos, Jaime Sevilla, Lennart Heim, Tamay Besiroglu, Marius Hobbhahn, Anson Ho, 'Will we run out of data? An analysis of the limits of scaling datasets in Machine Learning', arXiv, 26 October 2022, <https://arxiv.org/abs/2211.04325>

<sup>43</sup> 'FTC and DOJ Charge Amazon with Violating Children's Privacy Law by Keeping Kids' Alexa Voice Recordings Forever and Undermining Parents' Deletion Requests', Federal Trade Commission, 31 May 2023, <https://www.ftc.gov/news-events/news/press-releases/2023/05/ftc-doj-charge-amazon-violating-childrens-privacy-law-keeping-kids-alexa-voice-recordings-forever>; Elisa Jillson, 'Hey, Alexa! What are you doing with my data?', Federal Trade Commission, 13 June 2023, <https://www.ftc.gov/business-guidance/blog/2023/06/hey-alexa-what-are-you-doing-my-data>

## Suggested points for further consideration:

- 1. To address the most pressing concerns about current and future anti-competitive behavior in the AI market, the Federal Trade Commission should explore structural interventions targeted to the largest cloud infrastructure companies and their commercial AI model offerings.**

Cloud infrastructure providers are gatekeepers for significant amounts of economic activity across the entire market economy.<sup>44</sup> These gatekeeping concerns are particularly acute in the AI market, as cloud firms are in direct competition with the businesses that depend on them.<sup>45</sup> Moreover, the race to retain a competitive advantage in both cloud computing and large-scale AI may be driving cloud infrastructure firms to rush their commercial AI offerings to market prior to conducting adequate testing or validation, or consideration of potential harms to consumers and compliance with existing laws. Such behavior exemplifies what the antitrust scholars Ariel Ezrachi and Maurice Stucke have termed *toxic competition*.<sup>46</sup>

To address the most pressing concerns about current and future anti-competitive behavior in the AI market, the Federal Trade Commission should explore structural interventions that would address potentially unlawful conduct by these firms. It should also explore whether there are conflicts of interest that would impact cloud infrastructure providers' customer base more broadly, given potential limits to their computational bandwidth caused by these companies' large-scale AI offerings.

Additional concerns may be introduced within the compute market for AI more narrowly. Here, the FTC should explore the potential anti-competitive effects of tying compute software to chip design, such as architectures that link CUDA to the design of Nvidia's GPUs and TensorFlow to Google's TPU hardware.

The FTC's observations in the Nvidia-Arm merger case are germane to understanding the nature of vertical integration and concentration in the compute supply chain for AI.<sup>47</sup> Blocking this merger has already had positive effects on the market, averting an even greater level of

---

<sup>44</sup> Investigation of Competition in Digital Markets, Subcommittee on Antitrust, Commercial and Administrative Law of the Committee on the Judiciary of the House of Representatives, 117th Congress, July 2022, <https://www.govinfo.gov/content/pkg/CPRT-117HPRT47832/pdf/CPRT-117HPRT47832.pdf>

<sup>45</sup> Lina M. Khan, 'The Separation of Platforms and Commerce', Columbia Law Review 973, 2019, [https://scholarship.law.columbia.edu/faculty\\_scholarship/2789/](https://scholarship.law.columbia.edu/faculty_scholarship/2789/)

<sup>46</sup> Maurice E. Stucke, 'The Relationship Between Privacy and Antitrust', Notre Dame Law Review, Forthcoming, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4042262](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4042262)

<sup>47</sup> 'Statement Regarding Termination of Nvidia Corp.'s Attempted Acquisition of Arm Ltd.', Federal Trade Commission, 14 February 2022, <https://www.ftc.gov/news-events/news/press-releases/2022/02/statement-regarding-termination-nvidia-corps-attempted-acquisition-arm-ltd>

concentration in the compute industry by ensuring that no single company controls both CPUs (which are often based on Arm's designs) and GPUs (developed by Nvidia), essential resources for training and utilizing AI models. As noted by Bureau of Competition Director Holly Vedova, this was the first abandonment of a litigated vertical integration merger in years; the FTC should continue to monitor and intervene to prevent harms associated with vertical integration in AI markets through early stage intervention.

We expect structural separations to have the following effects:

- It is likely to address self-preferencing concerns among cloud providers and their AI offerings, and among chip designers and compute software.
- It eliminates the incentive to construct bundling and tying arrangements between compute and AI and among different layers of the compute stack.
- It reduces the drive to push AI products first to market, as well as toxic competition dynamics that are harmful to the profitability of businesses themselves.
- It would have the salutary effect of also addressing challenging accountability questions in the AI supply chain,<sup>48</sup> structuring the market in a manner that enables clearer distribution of responsibility and more effective regulation, such as mandated documentation tailored appropriately to the scope and context of use and pre-market testing of AI systems before they go to market to avert consumer harms.

## **2. The Federal Trade Commission should enforce the relevant antitrust laws against likely anti-competitive behavior, including restrictive licensing regimes and unconventional business arrangements designed to evade Hart-Scott-Rodino Act thresholds.**

The FTC should aggressively prosecute instances of anti-competitive conduct when they occur, preventing firms from further cementing their market dominance.<sup>49</sup> Examples of such conduct may include:

- Microsoft implemented a new policy when it launched the new version of its Bing search engine, triggering high fees for search engines that are customers of Bing should they develop their own alternatives to ChatGPT.<sup>50</sup>

---

<sup>48</sup> Jennifer Cobbe, Michael Veale, and Jatinder Singh, 'Understanding Accountability in Algorithmic Supply Chains', 2023 ACM Conference on Fairness, Accountability, and Transparency (FAcCT '23), 7 April 2023, <https://papers.ssrn.com/abstract=4430778>

<sup>49</sup> Sarah Myers West, 'Competition Authorities Need to Move Fast and Break up AI', Financial Times, 17 April 2023, <https://www.ft.com/content/638b5be7-fab7-4fe6-a0cf-7dabefcdd722>

<sup>50</sup> Tom Dotan and Deepa Seetharaman, 'The Awkward Partnership Leading the AI Boom', The Wall Street Journal, 13 June 2023, <https://www.wsj.com/articles/microsoft-and-openai-forge-awkward-partnership-as-techs-new-power-couple-3092de51>

- The unusual deal structure of Microsoft's arrangement with OpenAI may be designed to bypass the disclosure requirements of the Hart-Scott-Rodino Act.<sup>51</sup>
- Nvidia's provisioning of chips to ensure that firms that are not competing in chip development are the first recipients of H100s could be anti-competitive.<sup>52</sup>
- If the slowing of ChatGPT is related to Microsoft Azure's limited cloud bandwidth, this could constitute a degradation of product quality.

### **3. The FTC should clarify the legality of data use for model training, including whether the use of data by cloud providers to train their AI models violates consumer control over personal data and whether such data use should be considered in the FTC's commercial surveillance Advanced Notice of Proposed Rulemaking.**

Since data is an important differentiator for AI model quality, cloud providers might be engaging in anti-competitive behavior if they utilize user data to train AI models. They might also be violating well-established data protection principles like purpose limitation. Such potentially unilateral use of data by cloud providers to train AI models would also violate consumer control over personal data, a principle the FTC has outlined in its recent Amazon Alexa case.<sup>53</sup> The FTC should also consider such secondary use of commercial surveillance data as part of its Advanced Notice of Proposed Rulemaking on Consumer Surveillance and Data Security: this relates to data practices that potentially harm consumers and competition.<sup>54</sup>

---

<sup>51</sup> Jordan Novet, 'TECH Microsoft's \$13 billion bet on OpenAI carries huge potential along with plenty of uncertainty', CNBC, 8 April 2023, <https://www.cnn.com/2023/04/08/microsofts-complex-bet-on-openai-brings-potential-and-uncertainty.html>; Tammy Zhu, 'Microsoft's Quasi-acquisition of OpenAI: Clever Circumvention of Antitrust Scrutiny?', Tammy's Takes, 23 January 2023, <https://tammystakes.substack.com/p/microsofts-quasi-acquisition-of-openai>

<sup>52</sup> Anissa Gardizy, 'Why Nvidia Aids Cloud Rivals of AWS, Google and Microsoft', The Information, 1 June 2023, <https://www.theinformation.com/articles/why-nvidia-aids-cloud-rivals-of-aws-google-and-microsoft>

<sup>53</sup> 'FTC and DOJ Charge Amazon with Violating Children's Privacy Law by Keeping Kids' Alexa Voice Recordings Forever and Undermining Parents' Deletion Requests', Federal Trade Commission, 31 May 2023, <https://www.ftc.gov/news-events/news/press-releases/2023/05/ftc-doj-charge-amazon-violating-childrens-privacy-law-keeping-kids-alexa-voice-recordings-forever>

<sup>54</sup> 'Commercial Surveillance and Data Security Rulemaking', Federal Trade Commission, 11 August 2022, <https://www.ftc.gov/legal-library/browse/federal-register-notices/commercial-surveillance-data-security-rulemaking>