# Partnerships Between Cloud Service Providers and AI Developers

*FTC Staff Report on AI Partnerships & Investments 6(b) Study*


Federal Trade Commission
Office of Technology Staff
*January 2025*

# Table of Contents

# 1. Executive Summary

This report presents FTC Office of Technology ("OT") staff's findings from the agency's 6(b) study into three partnerships involving generative artificial intelligence ("AI"): Microsoft-OpenAI, Amazon-Anthropic, and Google-Anthropic.[1] These partnerships involve relationships between the world's current largest Cloud Service Providers ("CSPs") and two of the most prominent AI model developers. These partnerships therefore have potential for significant impact on AI technology, workers, and consumers. Through the 6(b) study, OT staff sought to better understand the nature of these relationships, including any ongoing obligations between the companies, to evaluate their potential impact on the marketplace.[2]

For an individual using an AI chatbot to write a wedding speech or a small business owner generating logo ideas with an AI image generation model, corporate partnerships may seem abstract or distant. Yet, these partnerships may potentially impact AI model development—including which firms may effectively participate in the marketplace—and may determine many aspects of those individuals' and firms' experiences. For these reasons, the agency's 6(b) study sought to understand the motivations and impacts of these partnerships on cloud computing and generative AI spaces.

The Commission issued the 6(b) orders to the five respondents in January 2024 and the respondents provided information through September 2024. AI technology is rapidly evolving, and therefore the context of this work may continue to evolve. The findings represent the information available to OT staff about these corporate partnerships as of September 2024 as well as publicly available information through January 2025.

This report is aimed at helping the agency, the public, and policymakers deepen their understanding of the corporate partnerships formed between the generative AI developers and CSPs included in this study. This report summarizes the findings from the 6(b) study, providing timely insight to the general public that is not otherwise available, including through previous reporting on the partnerships. It also equips the agency to better evaluate the attributes and potential implications of partnerships involving large technology companies and the impact they may have on hundreds of millions of consumers, countless businesses, and huge segments of the economy now, and not years after the fact. Even in places where public reporting had already described certain partnership features, non-public information helped to enrich staff's—and therefore this report's—analysis of the partnerships.

The information presented in this report has been aggregated or anonymized as necessary to protect trade secrets and confidential or privileged commercial or financial information submitted by respondents.

Staff notes that this report is not a formal legal or economic analysis of the partnerships or markets and should not be interpreted as such. The observations discussed in this report are limited to a narrow set of information and documents submitted by the five 6(b) respondents through September 2024. A comprehensive analysis of the subject partnerships is beyond the scope of the report. The technologies

[1] For the purposes of this report, generative AI is used to refer to the class of machine learning models that emulate the structure and characteristics of input data in order to generate derived synthetic content. This can include images, videos, audio, text, and other digital content, as defined in Exec. Order No. 14110, 88 Fed. Reg. 75191 (Nov. 1, 2023), https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence. This report uses the term AI as a shorthand for generative AI, because generative models are the only ones we discuss here except where otherwise noted.

[2] In this report we use the terms "marketplace" and "market" interchangeably. The use of these terms is meant to reference the word as it is generally understood; it should not be construed as an attempt to define an antitrust market.

discussed in this report are rapidly evolving, and therefore the implications of the partnerships may continue to evolve.

## Summary of Findings

The report contributes to the public's understanding of the evolving AI landscape in three ways. After providing background and summarizing the context in which this study was conducted, the report first describes the AI technology stack, according to OT staff's current understanding. Second, the report summarizes key terms of the AI partnerships based on review of the non-public documents produced in response to the study, as well as public reporting on, and public statements by, the respondents. Third, the report highlights certain areas to watch going forward: how the AI partnerships may affect access to certain inputs, how contractual terms and technical barriers may affect switching costs, and whether information exchanged in the partnerships might provide CSP partners access to sensitive technical and business information that may be unavailable to others. Since the study was limited to three corporate partnerships, any analysis of partnership terms in this report should not necessarily be construed to apply to all potential relationships between CSPs and AI developers.

Through information received via the 6(b) study, including non-public information, staff explored the potential "nature and extent of interactions with and influence over" each respondent from the partnership.[3]

### Key Terms of the AI Partnerships

Section 4 of this report highlights the following key terms[4] of the AI partnerships:

1. **The partnerships provide CSP partners significant equity and certain revenue-sharing rights in their AI developer partners.** The partnerships also leave open the possibility that the AI developers could be fully acquired by the CSPs in the future.
2. **The partnerships provide CSP partners certain consultation, control, and exclusivity rights with respect to their AI developer partners to varying degrees.** In some cases, the agreements include exclusive or preferential treatment of the CSP partner and the ability to potentially influence the AI model developer partner through board seats and consultation rights.
3. **The partnerships include cloud commitments that require AI developers to spend a large portion of their CSP partner's investment on cloud services from their partner.** This feature of the partnerships—circular spending—is one avenue through which CSP partners may potentially aim to reduce the magnitude of potential loss invested directly into their partners and cloud infrastructure to serve their partners.
4. **The partnerships permit the sharing of key resources and information.** These resources include, to varying degrees, discounted access to computing resources for AI developer partners; intellectual property, and performance and financial data for CSP partners; and engineering personnel and training data for both partners. The partnerships include co-development plans for CSP-designed semiconductor chips optimized for the needs of AI developer partners' models.

[3] Fed. Trade Comm'n, AI Investments and Partnerships 6(B) Study Sample Order (2024), https://www.ftc.gov/system/files/ftc_gov/pdf/P246201_AI_Investments_6%28b%29_Order_and_Resolution.pdf.
[4] Each of these terms may not apply to each of the partnerships; more specificity can be found in Section 4 of the report.

5. **The partnerships offer CSP and AI developer partners opportunities to expand current products.** In some cases, this includes integration into CSP partners' first-party products or deployment of AI developer partners' models on CSP partners' cloud platforms.

**Areas to Watch Regarding Potential Implications of the AI Partnerships**

Section 5 of this report identifies certain areas to watch going forward, which are not necessarily representative of all the partnerships' potential implications. A fulsome analysis of these issues is beyond the scope of this report but could be done using traditional antitrust tools and principles. The areas to watch include the following:

1. **The partnerships could affect access to certain inputs, such as computing resources and engineering talent.** The partnerships might affect the likelihood that a CSP partner would consider limiting access to inputs such as computing resources for AI developers other than its partner. The partnerships could also affect the availability of AI developer engineering talent. This could impact competition for both AI developer partners and non-partner AI developers.
2. **The partnerships could increase contractual and technical switching costs for AI developer partners.** The partnership agreements may increase switching costs via contractual commitments (including exclusivity requirements) and technical barriers. This can make it more difficult for AI developer partners to change CSPs or restrict their use of multiple CSPs.
3. **The partnerships provide CSP partners access to sensitive technical and business information that may be unavailable to others.** The partnership agreements provide CSPs access to certain technical and business information that other CSPs and AI developers may not have, including information related to generative AI models, AI development methods, confidential chip co-design, partner finances, and customer usage and revenue numbers. CSPs may be able to use this information to develop their own internal models and applications, including products that may compete with those of their partners.

# 2. Background

## 2.1 - Current Technological Landscape

The rapid and widespread development and deployment of generative AI has driven significant changes across society, including knowledge acquisition, creativity, and productivity. Generative AI has captured interest across commercial sectors—including in banking and financial services, healthcare, transportation, legal and professional services, customer service, travel, journalism, education, and marketing.

Although commercial generative AI is not completely new, public interest and investments in generative AI have sharply increased in recent years.[5] In 2023, billion-dollar AI partnerships began to form or grow, with the potential to impact the competitive dynamics of the space and therefore impact the quality and availability of AI products.

[5] *Generative AI Venture Capital Investment Globally on Track to Reach $12 Billion in 2024, Following Breakout Year in 2023,* EY IRELAND (May 16, 2024), https://www.ey.com/en_ie/news/2024/05/generative-ai-venture-capital-investment-globally-on-track-to-reach-12-billion-dollar-in-2024-following-breakout-year-in-2023.

Public reporting in late 2023 indicated that CSPs Microsoft, Alphabet (*i.e.*, Google),[6] and Amazon had invested billions of dollars in AI developers Anthropic and OpenAI, that some of the AI developers had committed to billions of dollars in cloud spending in return, and that the partnerships also involved other substantial non-monetary value exchanges.[7]

Public reporting and press releases by the partners have described the following partnership details in terms of monetary value:

*Table 1*

| | **Summary of Publicly Reported Investment $** | | |
|---|---|---|---|
| | **Microsoft-OpenAI** | **Amazon-Anthropic** | **Google-Anthropic** |
| **Investment amount** | $13.75 billion[8] | $8 billion[9] | $2.55 billion[10] |

However, publicly available information and the companies' marketing and press releases still left gaps in the public and staff's understanding of the nature of the partnerships—including both the potential market-level impact of the partnerships and more granular details of the exchange—which motivated this 6(b) study.

## 2.2 - Context of the 6(b) Study

In January 2024, the Commission voted to authorize a study pursuant to Section 6(b) of the FTC Act into the partnerships between two leading AI developers and three large CSPs.[11] The orders focused on the

[6] This report refers to Alphabet in this way because a number of Alphabet's services are Google-branded, such as Google Cloud Platform.
[7] Berber Jin & Tom Dotan, *Tech Giants Spend Billions on AI Startups—and Get Just as Much Back*, WALL STREET JOURNAL (Nov. 3, 2023), https://www.wsj.com/tech/ai/ai-deals-microsoft-google-amazon-7f624054.
[8] Microsoft has publicly stated that it invested $13 billion prior to September 30, 2024; public reporting suggests that Microsoft invested an additional $750 million in October 2024. *See* Berber Jin & Corrie Driebusch, *The $14 Billion Question Dividing OpenAI and Microsoft*, WALL STREET JOURNAL (Oct. 18, 2024), https://www.wsj.com/tech/ai/the-14-billion-question-dividing-openai-and-microsoft-71cf7d37 ("Microsoft has invested $13.75 billion in OpenAI since 2019, including its share of the startup's latest $6.6 billion fundraise."); Microsoft Corporation Form 10-Q for the Quarter Ended September 30, 2024, https://www.sec.gov/Archives/edgar/data/789019/000095017024118967/msft-20240930.htm ("We have an investment in OpenAI Global, LLC ('OpenAI') and have made total funding commitments of $13 billion.").
[9] Amazon Staff, *Amazon and Anthropic deepen their shared commitment to advancing generative AI*, AMAZON (Mar. 27, 2024), https://www.aboutamazon.com/news/company-news/amazon-anthropic-ai-investment; Amazon Staff, *Amazon and Anthropic deepen strategic collaboration*, AMAZON (Nov. 22, 2024), https://www.aboutamazon.com/news/aws/amazon-invests-additional-4-billion-anthropic-ai.
[10] Berber Jin & Miles Kruppa, *Google Commits $2 Billion in Funding to AI Startup Anthropic*, WALL STREET JOURNAL (Oct. 27, 2023), https://www.wsj.com/tech/ai/google-commits-2-billion-in-funding-to-ai-startup-anthropic-db4d4c50 ("Google agreed to invest up to $2 billion in Anthropic, building on its earlier investment in the artificial-intelligence company and adding fuel to the race between startups trying to achieve the next big breakthrough in the emerging technology. ... Google, owned by parent company Alphabet, already invested $550 million in Anthropic earlier this year.").
[11] Press Release, Fed. Trade Comm'n, FTC Launches Inquiry into Generative AI Investments and Partnerships (Jan. 25, 2024), https://www.ftc.gov/news-events/news/press-releases/2024/01/ftc-launches-inquiry-generative-ai-investments-partnerships. As discussed below, each of the CSP partners also develops and deploys generative AI products. However, their unique role in the partnerships is defined by their provision of cloud services, so this report refers to them as CSPs.

Microsoft-OpenAI, Amazon-Anthropic, and Alphabet-Anthropic partnerships because, based on public reporting, these were the largest such partnerships.[12]

The 6(b) orders included a targeted set of requests with the goal of developing a baseline understanding of the relationships between these CSPs and generative AI model developers. The orders sought non-public information to advance the agency's understanding of the potential impact of these partnerships on generative AI model development.

The Commission has used its Section 6 authority for decades to better understand market practices like these partnerships.[13] Section 6(b) empowers the Commission to require an entity to file "annual or special ... reports or answers in writing to specific questions" to provide information about the entity's "organization, business, conduct, practices, management, and relation to other corporations, partnerships, and individuals."[14] The Commission's 6(b) authority enables it to conduct wide-ranging studies that do not have a specific law enforcement purpose. Section 6(f) authorizes the Commission to "make public from time to time" portions of the information that it obtains, where disclosure would serve the public interest.[15]

The Commission is publicly reporting information collected from the 6(b) study in a manner consistent with Sections 6(f) and 21(d) of the FTC Act by aggregating or anonymizing as necessary to protect trade secrets and confidential or privileged commercial or financial information.

While some agency 6(b) studies seek to understand relatively stable market practices,[16] the various partnerships related to generative AI have continued to evolve and expand. Even in the short time since the Commission launched this 6(b) study, new relationships between CSPs and AI companies have been announced.[17]

Given the rapidly changing generative AI landscape, staff adopted a targeted approach to this study, carefully drafting the Special Orders to balance speed and timeliness with a desire to understand the significance of these quickly moving partnerships. The Commission requested information on five narrow specifications related to the partnerships between three major cloud service providers and two leading AI model developers. The summary of the specifications from the agency's press release[18] are as follows:

1. Information regarding the agreements themselves and related documents that explain the strategic rationale of the investment/partnership;
2. The practical implications of a specific partnership or investment, including decisions around new product releases, governance or oversight rights, and the topic of regular meetings;

---

[12] Chris Metinko, *The Biggest Of The Big: AI Startups Raised Huge — These Were The Largest Deals Of 2023*, CRUNCHBASE NEWS (Dec. 21, 2023), https://news.crunchbase.com/ai/biggest-ai-startups-openai-msft-eoy-2023/.
[13] *See e.g.,* Office of Policy Planning, Fed. Trade Comm'n, History of Section 6 Report-Writing at the Federal Trade Commission (Apr. 1981), https://www.ftc.gov/reports/history-section-6-report-writing-federal-trade-commission.
[14] 15 U.S.C. § 46(b); *see also* Fed. Trade Comm'n, A Brief Overview of the Federal Trade Commission's Investigative, Law Enforcement, and Rulemaking Authority (May 2021), https://www.ftc.gov/about-ftc/mission/enforcement-authority.
[15] 15 U.S.C. § 46(f).
[16] The Commission's tobacco 6(b) studies, for example, have spanned decades.
[17] For instance, Amazon invested an additional $4 billion in Anthropic in November 2024 and announced that AWS would be Anthropic's "primary training partner" (in addition to its "primary cloud provider") and that Anthropic would be using Amazon's Trainium and Inferentia chips "to train and deploy its future foundation models." *See* Amazon Staff, *supra* note 9; *see also* deals involving Inflection (with Microsoft), Character.ai (with Google), Adept (with Amazon), and Mistral (with Microsoft).
[18] Fed. Trade Comm'n, *supra* note 11.

3. Analysis of the transactions' competitive impact, including information related to market share, competition, competitors, markets, potential for sales growth, or expansion into product or geographic markets;
4. Competition for AI inputs and resources, including the competitive dynamics regarding key products and services needed for generative AI; and
5. Information provided to any other government entity, including foreign government entities, in connection with any investigation, request for information, or other inquiry related to these topics.

The full order specifications can be found in Appendix A.

## 2.3 - Technology Acquisitions and Partnerships

Parts of the CSPs' current strategies have their roots in past acquisitions. For example, Google's 2014 acquisition of DeepMind for reportedly over $500 million contributed to Google's AI efforts today.[19] DeepMind's former CEO, Demis Hassabis, is now the CEO of Google DeepMind.[20] Google's flagship generative AI models, the Gemini family of models, are built by this team.[21] Similarly, Amazon made an important acquisition in 2015, purchasing Annapurna Labs for a reported $370 million.[22] Annapurna Labs designs custom chips for Amazon, including those designed specifically for training and deploying AI models.[23] These custom chips are an essential part of Amazon's cloud and AI strategies—Amazon has even gone so far as to describe Annapurna as "the 'secret sauce' behind AWS's success."[24]

The Hart-Scott-Rodino ("HSR") Act,[25] which amended the Clayton Act in 1976, requires companies to file premerger notifications with the Federal Trade Commission and the Antitrust Division of the Justice Department for certain large mergers and acquisitions before they occur.[26] That premerger notification process involves completing an HSR Form, also called a "Notification and Report Form for Certain Mergers and Acquisitions," with information about each company's business.[27]

In 2021, the FTC published a report that focused on nearly a decade of unreported acquisitions by the biggest technology companies—*i.e.*, Alphabet/Google, Amazon, Apple, Facebook, and Microsoft.[28] The

---

[19] Catherine Shu, *Google Acquires Artificial Intelligence Startup DeepMind for More Than $500M*, TECHCRUNCH (Jan. 26, 2014), https://techcrunch.com/2014/01/26/google-deepmind/.
[20] Sundar Pichai and Demis Hassabis, *Our next-generation model: Gemini 1.5,* GOOGLE (Feb. 15, 2024), https://blog.google/technology/ai/google-gemini-next-generation-model-february-2024/.
[21] *Id.*
[22] *Amazon to buy Israeli start-up Annapurna Labs*, REUTERS (Jan. 22, 2015), https://www.reuters.com/article/technology/amazon-to-buy-israeli-start-up-annapurna-labs-idUSKBN0KV0SG/.
[23] *How silicon innovation became the 'secret sauce' behind AWS's success,* AMAZON (July 27, 2022), https://www.amazon.science/how-silicon-innovation-became-the-secret-sauce-behind-awss-success ("Those products, and silicon innovations, have done a lot of talking since 2015, as the acquisition has led to, among other advancements, the development of five generations of the AWS Nitro System, three generations (1, 2, 3) of custom-designed, Arm-based Graviton processors that support data-intensive workloads, *as well as AWS Trainium, and AWS Inferentia chips* optimized for machine learning training and inference.") (emphasis added).
[24] *Id.*
[25] 15 U.S.C. § 18a; *see also* Fed. Trade Comm'n, Hart-Scott-Rodino Antitrust Improvements Act of 1976, https://www.ftc.gov/legal-library/browse/statutes/hart-scott-rodino-antitrust-improvements-act-1976.
[26] Fed. Trade Comm'n, Premerger Notification Program, https://www.ftc.gov/enforcement/premerger-notification-program.
[27] *Id.*
[28] Press Release, Fed. Trade Comm'n, FTC Staff Presents Report on Nearly a Decade of Unreported Acquisitions by the Biggest Technology Companies (Sept. 15, 2021), https://www.ftc.gov/news-events/news/press-releases/2021/09/ftc-staff-presents-report-nearly-decade-unreported-acquisitions-biggest-technology-companies.

inquiry was initiated under Chair Joe Simons in 2020,[29] and the resulting report made public by the Commission with a 5-0 vote.[30] Upon its publication, Chair Lina Khan remarked how this study underscores the "need for us to closely examine reporting requirements ... and to identify areas where the FTC may have created loopholes that are unjustifiably enabling deals to fly under the radar."[31] According to Commissioner Rebecca Slaughter, this study "[helped the Commission] understand the bigger picture patterns among the largest tech platforms" with regards to acquisitions that did not require ex-ante review.[32] The study found that the five companies had completed 616 non-HSR reportable transactions—94 exceeded the HSR "Size of Transaction" threshold, and 65 percent were between $1 million and $25 million.[33] Other notable findings include that less than two-thirds of the non-reported transactions involved the acquisition of domestic assets or firms and more than 75 percent of transactions included non-compete clauses for founders and key employees of the acquired entities.[34] Some of these findings have overlap with some of the findings staff highlight in this report related to AI.

Amid these varying acquisition strategies undertaken by large technology firms and the development of generative AI technology, new forms of agreements—AI partnerships—have emerged. The AI partnerships that are subject of the present 6(b) study may represent a new strategy by large technology firms seeking to form relationships with AI model developers and represent just one way that the largest CSPs have responded to the generative AI boom. For example, there has recently been a rise in agreements that reportedly include hiring a company's AI experts, licensing the company's AI technology, and paying a fee in exchange.[35] In addition to licensing technologies and hiring numerous employees, these deals have elevated former startup CEOs to some of the most important roles in AI at

---

[29] Press Release, Fed. Trade Comm'n, FTC to Examine Past Acquisitions by Large Technology Companies (Feb. 11, 2020), https://www.ftc.gov/news-events/news/press-releases/2020/02/ftc-examine-past-acquisitions-large-technology-companies.

[30] Fed. Trade Comm'n, *supra* note 28.

[31] Remarks of Chair Lina M. Khan Regarding Non-HSR Reported Acquisitions by Select Technology Platforms Commission File No. P201201 (Sept. 15, 2021), https://www.ftc.gov/system/files/documents/public_statements/1596332/remarks_of_chair_lina_m_khan_regarding_non-hsr_reported_acquisitions_by_select_technology_platforms.pdf; *see also* Fed. Trade Comm'n, Non-HSR Reported Acquisitions by Select Technology Platforms, 2010- 2019: An FTC Study, at 20 (2021) ("[N]ine additional transactions would have exceeded the HSR SOT threshold (*i.e.*, in addition to the 94 transactions already above the HSR SOT threshold) at the time of their consummation when adding the deferred or contingent compensation (that is separate, and in addition to their purchase price) to their purchase price.").

[32] Prepared Remarks of Commissioner Rebecca Kelly Slaughter Regarding Non-HSR Reported Acquisitions by Select Technology Platforms, 2010-2019: An FTC Study (Sept. 15, 2021), https://www.ftc.gov/system/files/documents/public_statements/1596324/rks_statement_on_tech_6b_91521.pdf.

[33] *See* Fed. Trade Comm'n, *supra* note 31, at 13. Transactions that exceed the size threshold must be reported unless certain other criteria are not met or statutory or regulatory exemptions apply.

[34] *Id.* at 16, 21; *see also* Fed. Trade Comm'n, *supra* note 28.

[35] Erin Griffith & Cade Metz, *The New A.I. Deal: Buy Everything but the Company*, NEW YORK TIMES (Aug. 8, 2024), https://www.nytimes.com/2024/08/08/technology/ai-start-ups-google-microsoft-amazon.html ("The deal was one of several unusual transactions that have recently emerged in Silicon Valley. While big tech companies typically buy start-ups outright, they have turned to a more complicated deal structure for young A.I. companies. It involves licensing the technology and hiring the top employees — effectively swallowing the start-up and its main assets — without becoming the owner of the firm."); *id.* (describing Microsoft's March 2024 deal with Inflection); Annie Palmer, *Amazon beefs up AI development, hiring execs from startup Adept and licensing its technology*, CNBC (Jun. 28, 2024), https://www.cnbc.com/2024/06/28/amazon-hires-execs-from-ai-startup-adept-and-licenses-its-technology.html (describing Amazon's June 2024 deal with Adept); Kalley Huang, Natasha Mascarenhas, & Stephanie Palazzolo, *Google Hires Character.AI Cofounders and Licenses Its Models*, THE INFORMATION (Aug. 2, 2024), https://www.theinformation.com/articles/google-hires-character-ai-cofounders-and-licenses-its-models (describing Google's August 2024 deal with Character.AI).

the CSPs—leading one venture capital ("VC") firm to refer to these deals as "Big Tech consolidating AI talent."[36]

Many of FTC's international counterparts are also working to examine and better understand these partnerships. The following are some examples of these efforts:

- In late 2023, the UK's Competition and Market Authority ("CMA") opened an invitation to comment—an initial information gathering step in advance of the launch of any formal investigation—into the partnership between Microsoft and OpenAI.[37]
- In January of 2024, the European Commission announced that it is looking into agreements between "large digital market players and generative AI developers and providers" and specifically whether "Microsoft's investment in OpenAI might be reviewable under EU Merger Regulation."[38]
- In February of 2024, the French Competition Authority opened inquiries *ex officio* to analyze the generative AI market and "examine these types of investments" such as Microsoft's investment into OpenAI and Amazon and Google's investments into Anthropic.[39]
- Most recently, the CMA launched a merger inquiry regarding Microsoft's transaction with Inflection AI in July 2024,[40] an investigation of Amazon's partnership with Anthropic in August,[41] and an investigation into the Google-Anthropic partnership in October.[42] The CMA has since closed these investigations.[43]

# 3. The AI Technology Stack

As with all complex software products, AI development and commercialization combine several different layers into a composite product or service. The partnerships at the center of the 6(b) study involve collaboration and exchange across these layers of the tech stack, from chip development to deployment of AI-powered applications. Therefore, to understand the partnerships at hand it is essential to understand these layers and how they function.

This section explains the following layers of the tech stack that CSPs and AI model developers contribute to, depend on, and benefit from: AI Semiconductor Chips, Cloud Computing, Data and Models, and AI

---

[36] *AI eating software*, ACCEL (2024) at 32, https://cdn.prod.website-files.com/6643a08d305ab77f8c7566b6/670f22a19ea69a94f9710c1a_16%20October%20-%20Accel%202024%20Euroscape.pdf.
[37] Press Release, CMA, CMA seeks views on Microsoft's partnership with OpenAI (Dec. 8, 2023), https://www.gov.uk/government/news/cma-seeks-views-on-microsofts-partnership-with-openai.
[38] Press Release, European Commission, Commission launches calls for contributions on competition in virtual worlds and generative AI (Jan. 8, 2024), https://ec.europa.eu/commission/presscorner/detail/en/ip_24_85.
[39] Press Release, Autorité de la Concurrence, Generative artificial intelligence: the Autorité starts inquiries ex officio and launches a public consultation open until Friday, 22 March (Feb. 8, 2024), https://www.autoritedelaconcurrence.fr/en/press-release/generative-artificial-intelligence-autorite-starts-inquiries-ex-officio-and-launches.
[40] *Microsoft / Inflection inquiry*, CMA (Apr. 24, 2024), https://www.gov.uk/cma-cases/microsoft-slash-inflection-ai-inquiry.
[41] *Amazon / Anthropic partnership merger inquiry*, CMA (Apr. 24, 2024), https://www.gov.uk/cma-cases/amazon-slash-anthropic-partnership-merger-inquiry.
[42] *Alphabet Inc. (Google LLC) / Anthropic merger inquiry,* CMA (July 30, 2024), https://www.gov.uk/cma-cases/alphabet-inc-google-llc-slash-anthropic-merger-inquiry.
[43] *Id. See also* CMA, *supra* note 40; CMA *supra* note 41.

Applications. The FTC has referred to this framework previously in its Technology Summit on AI.[44] While not exhaustive, this framework is meant to be illustrative of different technologies that are part of the development, deployment, and use of AI.

## 3.1 - AI Semiconductor Chips

Semiconductor chips are the hardware foundation for the digital world. They sit at the base of the technology stack, providing the computational resources to build and deploy other hardware and software applications.[45]

Specialized AI chips are used in training a large generative AI model.[46] After a model has been trained, specialized AI chips are then used to generate responses or other outputs in response to prompts—a process known as "inference."[47]

Specially designed chips for AI are therefore a key input into AI model training and deployment. In notable contrast to "general purpose" chips such as central processing units ("CPUs"), AI chips[48] are optimized for performing a large volume of parallel calculations, such as the ones required for training and inference.[49] AI chips include graphical processing units ("GPUs"), which were originally built for computer graphics processing but are now heavily used by AI developers;[50] and other chips designed specifically for training or running AI, such as Google's Tensor Processing Units ("TPUs").[51]

The CSP respondents in this study are each developing specialized chips for AI applications.[52] For example, Amazon designs chips for AI models via its Trainium and Inferentia series of chips;[53] Google has its TPUs;[54] and Microsoft is developing AI chips called Maia 100.[55] In contrast to GPUs designed by

---

[44] Fed. Trade Comm'n, FTC Tech Summit (Jan. 2024), https://www.ftc.gov/news-events/events/2024/01/ftc-tech-summit.

[45] Jake Siegel, *With a systems approach to chips, Microsoft aims to tailor everything 'from silicon to service' to meet AI demand*, MICROSOFT (Nov. 15, 2023), https://news.microsoft.com/source/features/ai/in-house-chips-silicon-to-service-to-meet-ai-demand/ ("Chips are the workhorses of the cloud. They command billions of transistors that process the vast streams of ones and zeros flowing through datacenters. That work ultimately allows you to do just about everything on your screen, from sending an email to generating an image in Bing with a simple sentence.").

[46] Saif M. Khan, *AI Chips: What They Are and Why They Matter*, CSET (Apr. 2020), https://cset.georgetown.edu/publication/ai-chips-what-they-are-and-why-they-matter/ ("Different types of AI chips are useful for different tasks. GPUs are most often used for initially developing and refining AI algorithms; this process is known as 'training.'").

[47] Kim Martineau, *What Is AI Inferencing?*, IBM (Oct. 5, 2023), https://research.ibm.com/blog/AI-inference-explained ("Inference is the process of running live data through a trained AI model to make a prediction or solve a task.").

[48] As used herein "AI chips" means chips with the capability to be used for AI training and inference and dedicated to AI usage. Certain AI chips, such as GPUs, may be used for other tasks.

[49] Andrew Lohn & Micah Musser, *AI and Compute*, CSET (Jan. 2022), https://cset.georgetown.edu/publication/ai-and-compute/.

[50] *What is a GPU?,* AWS, https://aws.amazon.com/what-is/gpu/.

[51] *Introduction to Cloud TPU*, GOOGLE CLOUD (Nov. 21, 2024), https://cloud.google.com/tpu/docs/intro-to-tpu.

[52] *See, e.g.,* Max. A. Cherney, *Amazon racing to develop AI chips cheaper, faster than Nvidia's, executives say*, REUTERS (July 25, 2024), https://www.reuters.com/technology/artificial-intelligence/amazon-racing-develop-ai-chips-cheaper-faster-than-nvidias-executives-say-2024-07-25/.

[53] *AWS Tranium,* AWS, https://aws.amazon.com/machine-learning/trainium/; *AWS Inferentia*, AWS, https://aws.amazon.com/machine-learning/inferentia/.

[54] GOOGLE CLOUD, *supra* note 51.

[55] Borkar et al., *Azure Maia for the era of AI: From silicon to software to systems*, MICROSOFT (Apr. 3, 2024), https://azure.microsoft.com/en-us/blog/azure-maia-for-the-era-of-ai-from-silicon-to-software-to-systems/.

other providers such as AMD and Nvidia, the physical AI chips designed by Amazon, Microsoft, and Google are not available to purchase—they are only available to access via those CSPs.[56]

Software to interact with AI chips is another essential part of the equation. To use AI chips, AI developers must rely upon software that is able to integrate with those chips. That software is not necessarily compatible with different types of AI chips. For example, Nvidia's software platform, CUDA, is compatible only with Nvidia chips.[57] CUDA is widely used, with Nvidia estimating that four million developers were using it as of May 2023.[58] Additionally, there are other software platforms for chips, including technology called OneAPI that was first developed by Intel[59] and an open-source project called Triton, first released by OpenAI in July of 2021.[60]

Chip fabrication and design are separate processes. Large chip designers like Nvidia and AMD outsource their chip manufacturing to companies like Taiwan Semiconductor Manufacturing Co. ("TSMC").[61]

## 3.2 - Cloud Computing

Cloud computing—often referred to as "cloud"—is defined by the National Institute of Standards and Technology as "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (*e.g.*, networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."[62] Companies across the economy use CSPs to access the servers, databases, networking, and other services needed to build and run their products and internal tools. These resources are generally available on-demand and can rapidly scale to meet the needs of a developer,[63] allowing for technical architecture and physical resources that may take months to build and provision on traditional or non-cloud infrastructure to be assembled in a matter of seconds in the cloud.[64]

---

[56] Matthew S. Smith, *Challengers Are Coming for Nvidia's Crown*, IEEE Spectrum (Sept. 16, 2024), https://spectrum.ieee.org/nvidia-ai ("None of these compete directly with Nvidia, as hyperscalers don't sell hardware to customers. But they do sell access to their hardware through cloud services, like Google's AI Hypercomputer, Amazon's AWS, and Microsoft's Azure.").
[57] *CUDA GPUs – Compute Capability,* NVIDIA, https://developer.nvidia.com/cuda-gpus.
[58] Catherine Shu & Rita Liao, *All the Nvidia news announced by Jensen Huang at Computex*, TechCrunch (May 28, 2023), https://techcrunch.com/2023/05/28/nvidia-computex-jensen-huang/ ("Nvidia Cuda [sic] computing model now serves four million developers and more than 3,000 applications. Cuda [sic] has seen 40 million downloads, including 25 million just last year alone.").
[59] *OneAPI*, INTEL, https://www.intel.com/content/www/us/en/developer/tools/oneapi/overview.html.
[60] *Introducing Triton: Open-source GPU programming for neural networks*, OPENAI (July 28, 2021), https://openai.com/index/triton/.
[61] *Nvidia CEO Reiterates Solid Partnership with TSMC,* PR NEWSWIRE (March 22, 2024), https://www.prnewswire.com/news-releases/nvidia-ceo-reiterates-solid-partnership-with-tsmc-302096782.html; *AMD, TSMC to produce new chips,* TAIPEI TIMES (June 4, 2024), https://www.taipeitimes.com/News/front/archives/2024/06/04/2003818833.
[62] Peter Mell & Timothy Grance, *The NIST Definition of Cloud Computing*, NAT. INST. OF STANDARDS AND TECHNOLOGY (Sept. 2011), https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800–145.pdf.
[63] *Id.* at 2 (listing "Essential Characteristics" including "On-demand self-service" and "Rapid elasticity").
[64] The FTC Office of Technology, Fed. Trade Comm'n, An Inquiry into Cloud Computing Business Practices: The Federal Trade Commission is seeking public comments, (Mar. 22, 2023), https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/03/inquiry-cloud-computing-business-practices-federal-trade-commission-seeking-public-comments.

Many sectors of the economy rely on CSPs, which therefore gives CSPs the potential for widespread impact.[65] For example, transportation,[66] financial systems,[67] and healthcare systems[68] all use CSPs to manage their infrastructure. Startups, across a range of sectors, are especially reliant on CSPs.[69]

Three of this study's respondents offer cloud services—Microsoft's Azure, Amazon's Amazon Web Services ("AWS"), and Alphabet's Google Cloud Platform ("GCP").

In the context of AI models, cloud providers' most relevant offering may be access to a large volume of servers with specialized AI chips that can be used to train and deploy models. The high demand for servers with AI chips has reportedly led to months-long waits for some AI developers for access to this type of hardware.[70]

However, the provision of cloud services for AI consists of more than just obtaining servers with specialized chips. It also requires data centers to house those chips, networking to connect them, cooling systems to handle the enormous amount of heat generated by them, and more.

Data centers, connected by intricate networks, form the core of cloud offerings. CSPs house their servers and other hardware in physical data centers located all around the world. Data centers are "complex infrastructure projects" that "require a lot of people and it's not just the servers that go into them, but also the construction effort, the planning, the sustainable siting of the land, the permitting and renovating," according to Microsoft's director of Azure global infrastructure, Alistair Speirs.[71]

Given the amount of data that is used to train and operate AI models, specialized networking needs to be installed through physical data centers to facilitate sending the massive amounts of data between servers. In particular, the widespread practice among large AI developers of parallelizing computations among multiple chips and quickly communicating the results of such computations between chips can mean that data center infrastructure is often specialized for AI needs.[72] Similarly, high-volume data transfers inherent to AI training and inference can require significant networking between data centers—an issue that was discussed in at least one partnership, according to internal documents.[73] AI applications also require substantial amounts of power.

---

[65] *Id.*
[66] *See, e.g., Parsons' Intelligent Traffic Solution Built on AWS Helps Cities Reduce Congestion and Cut Costs*, AWS, https://aws.amazon.com/partners/success/parsons/.
[67] Press Release, U.S. Dept. of the Treas., Treasury and the Financial Services Sector Coordinating Council Publish New Resources on Effective Practices for Secure Cloud Adoption (July 17, 2024), https://home.treasury.gov/news/press-releases/jy2467.
[68] U.S. Dept. of Health and Human Services, Guidance on HIPAA & Cloud Computing (Dec. 23, 2022), https://www.hhs.gov/hipaa/for-professionals/special-topics/health-information-technology/cloud-computing/index.html.
[69] *See, e.g.*, Engine's Comments on the Business Practices of Cloud Computing Providers, Docket FTC-2023-0028, https://www.regulations.gov/comment/FTC-2023-0028-0056 ("Unsurprisingly, cloud technologies were among the most widely used. For example, no startup captured by the survey was using its own hosting infrastructure.").
[70] Aaron Holmes & Anissa Gardizy, *AI Developers Stymied by Server Shortage at AWS, Microsoft, Google*, THE INFORMATION (Apr. 7, 2023), https://www.theinformation.com/articles/ai-developers-stymied-by-server-shortage-at-aws-microsoft-google?rc=lwb9yo.
[71] Matt Ashare, *Microsoft, playing the long game, invests billions globally to expand Azure empire*, CIO DIVE (June 10, 2024), https://www.ciodive.com/news/microsoft-cloud-azure-data-center-investments-generative-ai/717897/.
[72] *See, e.g.,* descriptions of the specialized networking components in supercomputers for each supercomputer profile listed here under "interconnect" here: TOP 500 (June 2024), https://www.top500.org/lists/top500/2024/06/.
[73] Respondent Document Submission [redacted]

Offering cloud services related to AI training and inference is an expensive, capital-intensive business, as evidenced by billions of dollars in spending from Microsoft, Amazon, and Alphabet in recent years. In Q4 of fiscal year 2024, Microsoft CFO Amy Hood stated that Microsoft's capital expenditures reached $19 billion for the quarter, with "Cloud and AI related spend represent[ing] nearly all of total capital expenditures" spread across building data centers, leasing data centers, and purchasing servers.[74] Likewise, AWS announced its capital expenditures reached $30.5 billion in the first half of 2024 and expects to spend more in the second half of 2024, the majority of which "will be to support the growing need for AWS infrastructure," particularly due to the "strong demand in both generative AI and our non-generative AI workloads."[75] Additionally, Alphabet announced in its Q2 2024 earnings call that its capital expenditures for the quarter were $13 billion, "driven overwhelmingly by investment in our technical infrastructure, with the largest component for servers, followed by data centers."[76]

## 3.3 - Data and Models

Generally, AI developers aim to build foundation models that are fast[77] and produce high quality responses while reducing harm and errors.[78] Developers may also require their model to function with text, audio, images, and to some extent video as input or output, the latter of which significantly increases computational intensity of the models.[79]

In order to achieve high levels of performance, AI developer partners need an extensive amount of computational power. Anthropic CEO Dario Amodei publicly stated in April of 2024 that then-current models cost close to $1 billion to train, and that "in 2025 and 2026, we'll get more towards $5 or $10 billion."[80] Having access to this amount of computational power may impact AI model development. As one CSP partner explained, "we have entered an AI training regime where the quality of results achievable across a broad range of AI tasks is mostly constrained by the amount of AI training computing resources that can be brought to bear on the training task."[81]

---

[74] Satya Nadella & Amy Hood, *Microsoft Fiscal Year 2024 Fourth Quarter Earnings Conference Call*, MICROSOFT (July 30, 2024), https://www.microsoft.com/en-us/Investor/events/FY-2024/earnings-fy-2024-q4.aspx ("Capital expenditures including finance leases were $19 billion, in line with expectations, and cash paid for P, P, and E was $13.9 billion. Cloud and AI related spend represents nearly all of total capital expenditures. Within that, roughly half is for infrastructure needs where we continue to build and lease datacenters that will support monetization over the next 15 years and beyond. The remaining cloud and AI related spend is primarily for servers, both CPUs and GPUs, to serve customers based on demand signals. For the full fiscal year, the mix of our cloud and AI related spend was similar to Q4.").

[75] Georgia Butler, *AWS revenue hits $26.3bn for Q2 2024, Amazon capex reaches $30.5bn*, DATA CENTER DYNAMICS (August 2, 2024), https://www.datacenterdynamics.com/en/news/aws-revenue-hits-263bn-for-q2-2024-amazon-capex-reaches-305bn/.

[76] *2024 Q2 Earnings Call*, ALPHABET, https://abc.xyz/2024-q2-earnings-call/.

[77] *I.e.,* minimize the amount of time between input of a prompt and model response.

[78] Wei-Lin Chiang et al.*, Chatbot Arena: An Open Platform for Evaluating LLMs by Human Preference*, ARXIV E-PRINTS (Mar. 7, 2024), https://arxiv.org/abs/2403.04132.

[79] *Hello GPT-4o*, OPENAI (May 13, 2024), https://openai.com/index/hello-gpt-4o/.

[80] *Transcript: Ezra Klein Interviews Dario Amodei*, NEW YORK TIMES (April 12, 2024), https://www.nytimes.com/2024/04/12/podcasts/transcript-ezra-klein-interviews-dario-amodei.html ("The models that are in training now and that will come out at various times later this year or early next year are closer in cost to $1 billion. So that's already happening. And then I think in 2025 and 2026, we'll get more towards $5 or $10 billion.").

[81] Respondent Document Submission [redacted]

In addition to this computational power, training data is also a key input when creating a generative AI model. Generative AI models are trained on massive amounts of information. Training data can be split up into roughly two types of data—data for pretraining and data for "fine-tuning" the model.[82]

Pretraining data consists of enormous datasets of examples of content. For example, Meta's Llama 3 is an example of a text-based generative AI model, also known as a large language model ("LLM"). Llama 3 was trained on "over 15 [trillion] tokens."[83] As a rule of thumb, one token is roughly ¾ of a word, according to OpenAI.[84] Thus, 15 trillion tokens is equivalent to roughly 11.25 trillion words, or roughly 112 million novels.[85]

Gathering, cleaning, and preparing this information is a major undertaking that may require its own research and development and may also be expensive, depending on the data sources. Commonly reported sources of pretraining data include data that has been scraped,[86] licensed from rightsholders,[87] or obtained from pre-existing services of the generative AI model developer that collect the data type in question (*e.g.*, photos and text from public posts on Facebook and Instagram).[88]

Fine-tuning data is used to create models that are better "aligned" with human preferences. One example of fine-tuning is a technique known as reinforcement learning from human feedback ("RLHF"),[89] in which humans rank a set of outputs from a model. Chatbots that are built on top of large language models may implement RLHF in order to better "align" a chatbot's responses.[90] Because the data collected via ranking feedback is done by humans, gathering this data is labor-intensive and often outsourced through

---

[82] *AI Foundation Models Initial Report*, CMA (Sept. 18, 2023) at 11-12, https://assets.publishing.service.gov.uk/media/650449e86771b90014fdab4c/Full_Non-Confidential_Report_PDFA.pdf.

[83] *Introducing Meta Llama 3: The most capable openly available LLM to date*, META (Apr. 18, 2024), https://ai.meta.com/blog/meta-llama-3/.

[84] *What are tokens and how to count them?* OPENAI (2024), https://help.openai.com/en/articles/4936856-what-are-tokens-and-how-to-count-them.

[85] This is a rough estimate using a novel length of 100,000 words. *What's the perfect length for a book?* PENGUIN BOOKS LTD. (Sept. 23, 2020), https://www.penguin.co.uk/articles/2020/09/book-length-debate-fiction-long-novels.

[86] Lauren Leffer, *Your Personal Information Is Probably Being Used to Train Generative AI Models*, SCIENTIFIC AMERICAN (Oct. 19, 2023), https://www.scientificamerican.com/article/your-personal-information-is-probably-being-used-to-train-generative-ai-models/.

[87] *AI content licensing deals: Where OpenAI, Microsoft, Google, and others see opportunity*, CB INSIGHTS (July 19, 2024), https://www.cbinsights.com/research/ai-content-licensing-deals/.

[88] Mike Clark, *Privacy Matters: Meta's Generative AI Features,* META (Sept. 27, 2023), https://about.fb.com/news/2023/09/privacy-matters-metas-generative-ai-features/ ("Publicly shared posts from Instagram and Facebook—including photos and text—were part of the data used to train the generative AI models underlying the features we announced at Connect.").

[89] *See, e.g.,* Daniel M. Ziegler et al., *Fine-Tuning Language Models from Human Preferences*, ARXIV E-PRINTS (Jan. 8, 2020), https://arxiv.org/abs/1909.08593.

[90] *See, e.g.,* Zhiqing Sun et al., *Aligning Large Multimodal Models with Factually Augmented RLHF*, ARXIV E-PRINTS (Sept. 25, 2023), https://arxiv.org/abs/2309.14525.

dedicated contractors.[91] Furthermore, a large and more diverse amount of this human ranking data may be required in order to improve language model performance.[92]

AI developers are often looking for new sources of training data, particularly now that a large portion of the publicly available internet data has already been used to train models.[93] Publicly available data may increasingly consist of AI-generated content, and training AI models on content generated by models may have negative consequences on model quality.[94] Therefore, model developers may put a premium on high-quality human-generated content. Owners of large platforms that host user-generated content have access to that data and may use that data to train models.[95]

The public availability and transparency of an AI model's code and weights may vary. The term "open-weights [foundation] models" is generally used to describe generative AI models where the code to run the model and the "weights" (*i.e.*, the billions or trillions of numbers that determine an AI model's outputs) are made available to download to the public for free. The FTC has previously discussed open-weights foundation models.[96] As of the writing of this report, none of the respondents' flagship models are open-weights foundation models.[97]

As FTC staff has stated previously, "while open-source software (OSS) is well-defined, there is still an active dialogue around what 'open' and 'open-source' should mean in the emerging context of AI models—and it is important to understand the range of definitions when assessing the potential impacts."[98] Openness of models is a spectrum, which includes at a minimum the availability of model weights but can also include availability of the code used to train the model, descriptions of, or complete availability of, training and fine-tuning data, and legal restrictions or lack thereof on use, modification, and redistribution.[99] By contrast, "closed models" are models that are only available from companies through their products and services—such as an application programming interface ("API")[100]—and not

---

[91] *See, e.g., Evaluation*, SCALE AI (2024), https://scale.com/evaluation/model-developers. Staff note that non-generative AI models, too, can be highly dependent on labelling labor. *See, e.g.,* Todd Bishop, *How this startup used AI to keep raccoons from invading my house*, GEEKWIRE (Nov. 27, 2024), https://www.geekwire.com/2024/how-this-startup-used-ai-to-keep-raccoons-from-invading-my-house/.

[92] *See, e.g.,* Long Ouyang et al., *Training language models to follow instructions with human feedback,* NEURIPS (2022), https://proceedings.neurips.cc/paper_files/paper/2022/file/b1efde53be364a73914f58805a001731-Paper-Conference.pdf.

[93] Deepa Seetharaman, *For Data-Guzzling AI Companies, the Internet Is Too Small*, WALL STREET JOURNAL (Apr. 1, 2024), https://www.wsj.com/tech/ai/ai–training–data–synthetic–openai–anthropic–9230f8d8.

[94] Ilia Shumailov et al., *AI models collapse when trained on recursively generated data*, NATURE VOL. 631, (2024), https://www.nature.com/articles/s41586-024-07566-y.

[95] FTC Staff in the Office of Technology and the Division of Privacy and Identity Protection, Fed. Trade Comm'n, AI (and other) Companies: Quietly Changing Your Terms of Service Could Be Unfair or Deceptive (Feb. 13, 2024), https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2024/02/ai-other-companies-quietly-changing-your-terms-service-could-be-unfair-or-deceptive.

[96] Staff in the Office of Technology, Fed. Trade Comm'n, On Open-Weights Foundation Models (July 10, 2024), https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2024/07/open-weights-foundation-models.

[97] Alex Heath, *Meta releases the biggest and best open-source AI model yet*, THE VERGE (July 23, 2024), https://www.theverge.com/2024/7/23/24204055/meta-ai-llama-3-1-open-source-assistant-openai-chatgpt.

[98] Staff in the Office of Technology, *supra* note 96.

[99] *See, e.g.,* Irene Solaiman, *The Gradient of Generative AI Release: Methods and Considerations*, FACCT (2023), https://dl.acm.org/doi/10.1145/3593013.3593981. Models which are neither fully closed nor fully open might be dubbed "clopen" in reference to the use of this term for sets in point-set topology that are both closed and open.

[100] In other words, responses are accessible through "calls" to the provider's servers. *See, e.g., Application Programming Interface (API),* NIST COMPUTER SECURITY RESOURCE CENTER, https://csrc.nist.gov/glossary/term/application_programming_interface.

through sharing the model weights. As staff has previously stated, the degree to which a model is "open" can have implications for competition and consumer protection.[101]

## 3.4 - AI Applications

Consumers interface with AI through a number of mediums, including mobile phones, gaming devices, car driving and navigation systems, healthcare devices, and finance software tools. These interactions, which are often enabled through text, voice, or imagery, generally take place within applications ("apps"). These interfaces enable companies to either augment existing products and services with AI or offer completely new products and services to consumers.

Some apps provide users access to interact directly with generative AI models. Examples include general-purpose chatbots like ChatGPT[102] and image generation tools like Stable Diffusion.[103] These offerings exist in desktop and mobile devices, in apps and through browsers.

Companies also integrate generative AI into existing products. Examples include Google integrating responses in the form of "AI Overviews" into its search results page,[104] Amazon integrating a chatbot into its shopping app,[105] Snap using OpenAI's APIs to build its "My AI" feature,[106] and Microsoft integrating large language models into its Microsoft 365 apps via its various "Copilot" offerings.[107] In addition to consumer-facing products, businesses may integrate generative AI into their own internal tools. Examples of this internal usage include summarization of internal documents[108] and code generation.[109] Finally, generative AI may be used behind the scenes in ways that are opaque to end-users—such as in content moderation systems on digital platforms.[110]

AI developers may charge subscription fees to use their products. For standalone AI applications, some companies rely on a paid subscription model. In addition to offering free tiers, some companies—including OpenAI, Anthropic, and Midjourney—offer paid options that include additional capabilities, early previews of new products, or higher limits on usage.[111]

---

[101] Staff in the Office of Technology, *supra* note 96.
[102] *ChatGPT*, OPENAI, https://chatgpt.com/. The term chatbot here refers to a consumer-facing text- or voice-based generative AI model that is meant to respond directly to user queries and is not limited to a single domain such as customer service.
[103] *Image Models*, STABILITY.AI, https://stability.ai/stable-image.
[104] Liz Reid, *Generative AI in Search: Let Google do the searching for you*, GOOGLE (May 14, 2024), https://blog.google/products/search/generative-ai-google-search-may-2024/.
[105] Rajiv Mehta, *How customers are making more informed shopping decisions with Rufus, Amazon's generative AI-powered shopping assistant*, AMAZON (Sept. 18, 2024), https://www.aboutamazon.com/news/retail/how-to-use-amazon-rufus.
[106] *Staying Safe with My AI,* SNAPCHAT SUPPORT, https://help.snapchat.com/hc/en-us/articles/13889139811860-Staying-Safe-with-My-AI.
[107] *Microsoft 365 Copilot,* MICROSOFT ADOPTION (2024), https://adoption.microsoft.com/en-us/copilot/.
[108] *Jump Start Solution: Generative AI document summarization*, GOOGLE CLOUD (July 21, 2023), https://cloud.google.com/architecture/ai-ml/generative-ai-document-summarization.
[109] Juyong Jiang et al., *A Survey on Large Language Models for Code Generation*, ARXIV E-PRINTS (Nov. 10, 2024), https://arxiv.org/abs/2406.00515.
[110] *Using GPT-4 for content moderation*, OPENAI (August 15, 2023), https://openai.com/index/using-gpt-4-for-content-moderation/.
[111] *See, e.g.*, *What is Claude Pro?,* ANTHROPIC, https://support.anthropic.com/en/articles/8325606-what-is-claude-pro; *Introducing ChatGPT Plus,* OPENAI (Feb. 1, 2023), https://openai.com/index/chatgpt-plus/; *Subscription Plans*, MIDJOURNEY, https://docs.midjourney.com/docs/plans.

AI developers may offer API access or license the use of their models to other companies. For example, AI models can be incorporated via a licensing agreement into search engines,[112] software development tools,[113] and cloud offerings.[114] For API usage, companies using the AI models may pay based on how much data they send or receive back from the API, and the usage of the data that is sent to the AI service provider may be governed by the AI provider's policies and any business agreements.[115]

Alternatively, an application may be built on a cloud that provides access to AI models as one of its offerings, an option that some cloud providers describe as providing more flexibility and security.[116] All three CSP respondents offer these "Model-as-a-Service" ("MaaS") products, where customers can access a range of models via the CSP's platform: Amazon offers Bedrock,[117] Microsoft offers Azure AI Model Catalog,[118] and Google offers Vertex AI.[119] Developers who want to integrate AI into their product or service without having to train an AI model themselves often turn to MaaS offerings for access to such models. For example, DoorDash uses Bedrock to access AI models that power its voice AI assistant.[120]

Consumer-facing AI products may be monetized by displaying advertising in their interfaces. For example, Google's AI Overview product, powered by its Gemini LLM, recently started showing ads related to a user's query.[121] Similarly, Microsoft started including ads in its Bing Chat (later rebranded to "Copilot") interface in November of 2023.[122]

## 4. Key Terms of the AI Partnerships

This section describes key terms of the AI partnerships as observed in the 6(b) study, including: (1) equity and revenue-sharing rights; (2) certain consultation, control, and exclusivity rights; (3) cloud spend commitments; (4) sharing of key resources and information; and (5) the potential to expand current products and scale new offerings. Access to non-public documents through the 6(b) study helped expand staff's understanding of the partnerships and their exact terms, beyond what had been reported publicly.

---

[112] *Your Everyday AI Companion | Microsoft Copilot*, MICROSOFT, https://www.microsoft.com/en-us/bing?ep=466&form=MA13P3&es=31.
[113] *Your AI Pair Programmer*, GITHUB COPILOT, https://github.com/features/copilot.
[114] *Amazon Bedrock*, AWS, https://aws.amazon.com/bedrock/.
[115] *See, e.g.*, the pricing page of OpenAI's API offering: *Pricing*, OPENAI, https://openai.com/api/pricing/.
[116] For example, Microsoft touts its offering of OpenAI models via Azure as providing "the security and enterprise promise of Azure. … With Azure OpenAI, customers get the security capabilities of Microsoft Azure while running the same models as OpenAI. Azure OpenAI offers private networking, regional availability, and responsible AI content filtering." *What is Azure OpenAI Service?: Comparing Azure OpenAI and OpenAI*, MICROSOFT (Nov. 5, 2024, https://learn.microsoft.com/en-us/azure/ai-services/openai/overview#comparing-azure-openai-and-openai.
[117] AWS, *supra* note 114.
[118] *Azure AI Model Catalog*, MICROSOFT, https://azure.microsoft.com/en-us/products/ai-model-catalog.
[119] *Vertex AI Platform*, GOOGLE CLOUD, https://cloud.google.com/vertex-ai?hl=en.
[120] *Building a Generative AI Contact Center Solution for DoorDash Using Amazon Bedrock, Amazon Connect, and Anthropic's Claude*, AWS (2024), https://aws.amazon.com/solutions/case-studies/doordash-bedrock-case-study/?did=cr_card&trk=cr_card.
[121] Ron Amadeo, *Next up in Google's dramatic overhaul of search: AI Overview ads*, ARS TECHNICA (May 23, 2024), https://arstechnica.com/gadgets/2024/05/next-up-in-googles-dramatic-overhaul-of-search-ai-overview-ads/.
[122] Kya Sainsbury-Carter, *Transforming Search and Advertising with Generative AI*, MICROSOFT (Sept. 21, 2023), https://about.ads.microsoft.com/en/blog/post/september-2023/transforming-search-and-advertising-with-generative-ai. Bing Chat was later rebranded to simply "Copilot." *See* Tom Warren, *Microsoft rebrands Bing Chat to Copilot, to better compete with ChatGPT*, THE VERGE (Nov. 15, 2023), https://www.theverge.com/2023/11/15/23960517/microsoft-copilot-bing-chat-rebranding-chatgpt-ai.

## 4.1 - The partnerships provide CSP partners significant equity and certain revenue-sharing rights in their AI developer partners.

As a result of the CSPs' ownership interests in, or rights to profits from, the AI developer partner, the CSP partners might benefit financially if their AI developer partner's value increases.[123] Staff notes that profit allocated to Microsoft by OpenAI is reportedly currently capped at $92 billion.[124]

Public reporting has indicated that both Amazon and Google have non-voting shares in Anthropic.[125] Staff is currently not aware of a public disclosure regarding the nature of Microsoft's ownership in OpenAI.

Public reporting has also posited that each of the partnerships contain revenue-sharing agreements that provide income for CSPs based on the success of their partner's AI models, including sharing revenue from the model developer partner's own first-party API.[126] Internal documents confirmed that revenue sharing is part of multiple partnerships in various ways, including [redacted][127]

The 6(b) documents reviewed by staff demonstrated that at least some respondents recognized that the acquisition of the AI developer by its CSP partner was a possibility.[128] One respondent, for example, wrote in internal documents that [redacted]

Other documents also indicated that some respondents were less receptive to a CSP acquiring its AI developer partner.[130] AI developer partners and CSP partners have also considered extending their partnerships with further funding.[131]

---

[123] *See, e.g.*, Respondent Document Submission [redacted] for one company's discussion of potential investment outcomes.

[124] *Our structure*, OPENAI, https://openai.com/our-structure/; Jessica Mathews and Jeremy Kahn, *Inside the Structure of OpenAI's looming new investment from Microsoft and VCs,* FORTUNE (Jan. 11, 2023), https://fortune.com/2023/01/11/structure-openai-investment-microsoft/.

[125] Billy Perrigo, *How Anthropic Designed Itself to Avoid OpenAI's Mistakes*, TIME (May 30, 2024), https://time.com/6983420/anthropic-structure-openai-incentives/.

[126] Anissa Gardizy, *Amazon Discussing New Multibillion-Dollar Investment in Anthropic*, THE INFORMATION (November 7, 2024), https://www.theinformation.com/articles/amazon-discussing-new-multibillion-dollar-investment-in-anthropic; *see also* Anissa Gardizy et al., *OpenAI Leaders Say Microsoft Isn't Moving Fast Enough to Supply Servers*, THE INFORMATION (Oct. 8, 2024), https://www.theinformation.com/articles/openai-eases-away-from-microsoft-data-centers.

[127] Respondent Document Submission [redacted]

[128] Respondent Document Submission [redacted]

[129] Respondent Document Submission [redacted].

[130] Respondent Document Submission [redacted].

[131] *See, e.g.*, Respondent Document Submission [redacted].

## 4.2 - The partnerships provide CSP partners certain consultation, control, and exclusivity rights with respect to their AI developer partners to varying degrees.

At least one partnership provides the opportunity for CSP executive leadership to consult with or advise their partner companies.[132] Internal documents described that one partnership gave the CSP partner input on [redacted][133] and at least one partnership also gave the CSP partner advance notice of important decisions.[134] These rights may provide a potential way for CSP partners to influence the AI model developer partners. However, these rights vary in the extent to which they have been exercised and the potential they have for impacting the day-to-day operations of AI developer partners.

Control can also come in the form of who is on the company's board of directors. Board members have several key duties which vary from organization to organization, but generally include strategic oversight, planning, budget approval and fundraising, legal compliance, and decision-making on significant corporate changes.[135]

For example, Microsoft previously occupied an observer seat on the board of OpenAI.[136] While it relinquished that seat in 2024, a company spokesperson also noted that Microsoft will continue to receive information from OpenAI and that they will "host regular stakeholder meetings to share progress on our mission and ensure stronger collaboration across safety and security."[137] In addition, one board member served on both Microsoft's and OpenAI's board contemporaneously for over three years.[138]

CSP partners also have certain exclusivity or parity rights through the partnerships—such as access to certain products, distribution capabilities, or brand association.

[132] *See, e.g.*, Respondent Document Submission [redacted]

[133] Respondent Document Submission [redacted]

[134] [redacted]

[135] *See, e.g.*, Kay Brancato et al., *The Role of U.S. Corporate Boards in Enterprise Risk Management*, THE CONFERENCE BOARD RESEARCH REPORT NO. R-1390-06-RR (Nov. 1, 2006), https://www.conference-board.org/publications/publicationdetail.cfm?publicationid=1190.

[136] Foo Yun Chee, *Microsoft ditches OpenAI board observer seat to stave off antitrust scrutiny*, REUTERS (July 10, 2024), https://www.reuters.com/technology/microsoft-ditches-openai-board-observer-seat-amid-regulatory-scrutiny-2024-07-10/.

[137] Mauro Orru and Christian Moess Laursen, *Microsoft Quits OpenAI's Board Amid Antitrust Scrutiny*, WALL STREET JOURNAL (July 10, 2024), https://www.wsj.com/tech/ai/microsoft-withdraws-from-openais-board-amid-antitrust-scrutiny-aab6ff1e.

[138] Kif Leswing, *Reid Hoffman steps down from OpenAI board to avoid potential conflicts of interest*, CNBC (Mar. 3, 2023), https://www.cnbc.com/2023/03/03/reid-hoffman-steps-down-from-openai-board-to-avoid-potential-conflicts.html.

In some cases, the agreements include exclusive or preferential treatment to the CSP partner,[139] or require parity with other cloud partners or the AI developer's own products.[140] Staff also notes that although the existence of exclusivity, preference, or parity provisions varied by partnership, in general, multiple partnerships have provisions that impose conditions or restrictions on the ability of AI developer partners to share accrued benefits with other non-partner cloud providers or companies.[141]

One CSP described the different ways in which its partnership could potentially provide parity as follows:

> "We've sought parity in multiple ways: Parity of making the models and updates of the same available at the same time as they do [redacted] [core ask]. Parity on features and functionality including performance, reliability, support, terms of service, regional availability, etc. Parity on pricing [redacted] vs [redacted] (they set the pricing). Parity on compensation for their sellers on sales on [redacted] (as compared to [redacted]). Parity on marketing (promotion of [redacted] as with [redacted])."[142]

In at least one case, a partnership effectively required the AI developer partner to keep the models on its partner's cloud service at parity with the models on the AI developer's first-party API.[143] In other words, the AI developer would not be able to release a new model on its own before launching it on its partner's cloud service. Nor would the AI developer partner be able to offer substantially better models to business or individual consumers as an incentive to use its models through a first-party API rather than through its partner's cloud.

### 4.3 - The partnerships include cloud commitments that require AI developers to spend a large portion of their CSP partner's investment on cloud services from their partners.

According to public reporting summarized in Table 2, the AI developer partners committed to spend billions of dollars on cloud services from their partners.[144] The committed spending on cloud computing means that, to a significant extent, the CSP is effectively investing in the partnership through the provision of cloud computing. This feature of the partnerships—circular spending—is one avenue

[139] *See, e.g.,* Respondent Document Submission [redacted].

[140] *See, e.g.*, Respondent Document Submission [redacted].

[141] *See* [redacted] *ee also* [redacted] Respondent Document Submission [redacted].

[142] Respondent Document Submission [redacted].

[143] *See, e.g.,* Respondent Document Submission [redacted], stating that [redacted] *See also* Respondent Document Submission [redacted], citing [redacted] *See* Respondent Document Submission [redacted].

[144] *See* Jin & Dotan, *supra* note 7.

through which CSP partners may have potentially aimed to reduce the magnitude of potential loss invested directly into their partners and cloud infrastructure to serve their partners.

*Table 2*

| | **Summary of Publicly Reported Spending Commitment $** | | |
|---|---|---|---|
| **Cloud spending commitment by AI developer** | **Microsoft-OpenAI** | **Amazon-Anthropic** | **Google-Anthropic** |
| | "Billions"[145] | At least $4 billion, not including commit after latest $4 billion investment[146] | At least $3 billion[147] |

AI developers' committed funds requirements varied by partnership, but funds can be spent on a range of computational tasks needed by AI developers, including training or fine-tuning foundation models; inference, *i.e.*, computing responses to consumer prompts on the AI developers' own products or via CSP model platforms; or storage services, such as storing training data or prompts.[148]

For Microsoft and OpenAI specifically, according to publicly disclosed information, OpenAI has been using Microsoft as its "exclusive cloud provider" dating since the start of their partnership in 2019, which was reaffirmed in 2023.[149]

At least one CSP order recipient claimed publicly that cloud spending commitments are separate from investment deals, and explained, for example, that the decision to invest in its developer partner was made by a separate team than the team negotiating any cloud spending commitment agreements.[150] However, 6(b) documents suggest that multiple respondents may view investments as linked to cloud spending commitments.[151] For example, in a list of strategic rationales for the partnership, one CSP listed first that the partnerships provided a "[l]arge [cloud service] [redacted] commit opportunity."[152]

One CSP noted that the cloud spend commitment had limited downside since any AI infrastructure built for its AI partner could be resold or used internally: "If [redacted] cannot raise/generate more funds, we

[145] *See* Jin & Dotan, *supra* note 7.
[146] *Id.*; *See also* Amazon Staff, *supra* note 9.
[147] *Id.*
[148] Respondent Document Submission [redacted] Respondent Document Submission [redacted]
[149] *Microsoft invests in and partners with OpenAI to support us building beneficial AGI*, OPENAI (July 22, 2019), https://openai.com/index/microsoft-invests-in-and-partners-with-openai/ ("We'll jointly develop new Azure AI supercomputing technologies, and Microsoft will become our exclusive cloud provider—so we'll be working hard together to further extend Microsoft Azure's capabilities in large-scale AI systems."); *Microsoft and OpenAI extend partnership*, MICROSOFT (Jan. 23, 2023), https://blogs.microsoft.com/blog/2023/01/23/microsoftandopenaiextendpartnership/ ("Exclusive cloud provider - As OpenAI's exclusive cloud provider, Azure will power all OpenAI workloads across research, products and API services.").
[150] *See* Jin & Dotan, *supra* note 7.
[151] Respondent Document Submission [redacted] Respondent Document Submission [redacted]
[152] Respondent Document Submission [redacted]

will at least make back our [capital expenditure] by [redacted], assuming [redacted]. In the improbable case of default, we will make back our Capex and have [redacted] at hand to resell to others or consume internally."[153]

## 4.4 - The partnerships permit the sharing of key resources and information.

The partnerships provide the partners with access to key resources, including access to computing resources, intellectual property, key personnel, performance and financial data, training data, and chip co-development opportunities to varying degrees. This section discusses each of these in turn.

### 4.4.1 - The partnerships provide AI developer partners, which require large amounts of computing resources to train and operate their models, discounted access to computing resources.

The partnerships provide AI developer partners with access to needed training and inference computing resources to train and operate their models. As explained in Section 3.3 AI developer partners need an extensive amount of computational power to create and deploy their models.

At least one respondent indicated in internal documents that it believes only a small number of players can scale past the capabilities of current state-of-the-art models, and that partnerships between AI developers and CSPs are an important way for model developers to obtain enough computing resources to keep ahead of the pack. [redacted]

Staff notes that the CSPs considered here offer certain discounts—such as committed use discounts—to the public, as well as private negotiated discounts to other customers. Some CSPs may publicly offer committed use discounts for a range of services, including data storage and specific resources needed to train and deploy AI models. For example, Google publicly offers discounts on certain GPU-based, specialized AI offerings in the United States of around 56% over three years.[155] Separately, one publicly available industry guide suggests that a 6 percent discount is reasonable on a $1 million annual committed spend on AWS.[156] The exact discount rates that CSPs offer via private negotiations and how the discounts are calculated, however, is not public.

Documents provided by respondents indicated that some discounts provided by CSP partners to their AI partners were significantly larger than certain publicly advertised discounts.[157] However, because the exact discount rates that CSPs offer outside of the partnerships was not a subject of the study, it is unclear how these rates compare to other privately negotiated discounts.

In some cases, provision of computing resources to AI developer partners went beyond funding and discounts. For instance, the Microsoft and OpenAI partnership specifically contains publicly disclosed

[153] Respondent Document Submission [redacted]
[154] Respondent Document Submission [redacted] Respondent Document Submission [redacted]
[155] *See VM Instance Pricing*, GOOGLE CLOUD, https://cloud.google.com/compute/vm-instance-pricing#committed_use. This is calculated based on a3-highgpu-1g instances in Iowa.
[156] Tony Chan, *AWS EDP 2024 Negotiation Guide*, CLOUDFORECAST.IO (May 24, 2024), https://www.cloudforecast.io/blog/aws-edp-guide/.
[157] Respondent Document Submission [redacted]

exclusivity, with Microsoft building a supercomputer "in collaboration with and exclusively for OpenAI."[158] The supercomputer— "one of the top five publicly disclosed supercomputers in the world," and in 2020 "the largest supercomputer ever built in the public cloud" according to Microsoft—provides OpenAI with exclusive access to this computing resource, at least during the period of exclusivity.[159]

### 4.4.2 - The partnerships provide CSP partners, which are developing their own AI models, access to assets and intellectual property related to the AI developers' cutting-edge models.

Documents illustrated that the partnerships provided the CSP partners with significant access to assets, IP, and research developed by the AI developer partners or co-developed between the two parties. At least one CSP viewed its partnership as a "bridge" to allow it to learn about frontier model development and to become a significant player in the foundation model space itself.[160] In other documents, at least one CSP saw its partnership as an alternative to internal model development efforts, with one CSP's executives agreeing in a meeting that "double-spend[ing] in this [generative AI] space" did not make sense given the CSP's investment in its AI developer partner.[161]

Generally, AI models themselves, training methods, and related source code are valuable intellectual property owned by the AI developers.[162] It is therefore notable that multiple partnerships provided CSP partners access to some of these core technologies, with rights to test or commercialize them in certain cases.

The 6(b) documents also highlighted that the partnerships would give the CSP partners direct access to cutting-edge model specifications,[163] development methodologies, and processes to test and deploy them in practice.[164] In the early stages of one partnership, one CSP partner noted that "potential collaborations - for exploration" included sharing "pretraining algorithmic insights" and "partner models available for [redacted] research."[165] Partner CSPs may seek access to this information to potentially help them

---

[158] Jennifer Langston, *Microsoft announces new supercomputer, lays out vision for future AI work*, MICROSOFT (May 19, 2020), https://news.microsoft.com/source/features/ai/openai-azure-supercomputer/ ("Built in collaboration with and exclusively for OpenAI, the supercomputer hosted in Azure was designed specifically to train that company's AI models.").

[159] *Id.* ("Microsoft has built one of the top five publicly disclosed supercomputers in the world, making new infrastructure available in Azure to train extremely large artificial intelligence models, the company is announcing at its Build developers conference."); Zachary Cavanell, *What runs ChatGPT? Inside Microsoft's AI supercomputer*, MICROSOFT MECHANICS BLOG (May 24, 2023), https://techcommunity.microsoft.com/t5/microsoft-mechanics-blog/what-runs-chatgpt-inside-microsoft-s-ai-supercomputer-featuring/ba-p/3830281 ("By the way, that system that we built back in 2020, was the fifth largest supercomputer in the world at the time and the largest supercomputer ever built in the public cloud. It's really the result of a collaborative effort between Microsoft engineering, Microsoft Research, the OpenAI organization, and NVIDIA.").

[160] Respondent Document Submission [redacted] ("To that end, we have been exploring strategic partnerships that would serve as a bridge until our own model is competitive with what we [redacted] will be.").

[161] Respondent Document Submission [redacted]

[162] *See, e.g.*, *Frontier Model Security*, ANTHROPIC (July 25, 2023), https://www.anthropic.com/news/frontier-model-security.

[163] Specifications refer to a set of descriptions and requirements for a software product. For an example of specifications in the generative AI context, *see, e.g., Model Spec,* OPENAI (May 8, 2024), https://cdn.openai.com/spec/model-spec-2024-05-08.html#overview.

[164] *See, e.g.*, [redacted]

[165] Respondent Document Submission [redacted].

develop their own models and improve their internal products and services. For example, more efficient model architectures can improve the cost-quality balance for LLMs.[166]

As noted in public reporting, Microsoft CEO Satya Nadella remarked, "If OpenAI disappeared tomorrow ... we have all of the rights to continue the innovation. Not just to serve the product, but we can go and just do what we were doing in partnership ourselves. We have the people, we have the compute, we have the data, we have everything."[167]

### 4.4.3 - The partnerships may permit CSP and AI developer partners to embed engineers in one another's companies, allowing for information transfer about technology and intellectual property.

Because highly skilled labor is an important input for generative AI, CSPs appear to be using the partnerships to gain access to relevant technical talent.

At least one partnership provided for the exchange of employees from one partner company to another—as part of these exchanges, employees may embed within teams at the partner company, co-develop products, and spend the majority of their time working with the partner company.[168] One respondent described embedded employees as "the primary vehicle" for transferring information about IP and technology.[169]

Another document submitted pursuant to this study described an interest in having senior employees, including data scientists and engineers, working with the partner to speed up inference times, improve fine-tuning of models, support model training, research cutting-edge model architectures, and more.[170]

Similarly, AI developer partners secured access to CSP partner technical support and engineering talent on various issues that arise related to their CSP partner's platforms, which could make it easier to avoid potentially disruptive events such as server failures or common development speedbumps.

Due to the complexity of cutting-edge AI models, failures and incompatibilities can occur when they are deployed by customers.[171] Documents produced by respondents indicate that developers required technical support and met with their CSP partners on a regular basis to "resolve issues that [were]

---

[166] *See, e.g.,* Canwen Xu & Julian McAuley, *A Survey on Model Compression and Acceleration for Pretrained Language Models*, ARXIV E-PRINTS (Nov. 29, 2022), https://arxiv.org/abs/2202.07105 (for an explanation of generative AI compression and acceleration efforts). Recent models that reflect advances in efficiency include GPT 4o-mini, Claude 3.5 Sonnet, and Gemini 1.5 flash. For benchmark examples, *see Claude 3.5 Sonnet*, ANTHROPIC (June 20, 2024), https://www.anthropic.com/news/claude-3-5-sonnet.

[167] Intelligencer Staff, *Satya Nadella on Hiring the Most Powerful Man in AI*, INTELLIGENCER (Nov. 21, 2023), https://nymag.com/intelligencer/2023/11/on-with-kara-swisher-satya-nadella-on-hiring-sam-altman.html.

[168] Respondent Document Submission [redacted].

[169] Respondent Document Submission [redacted].

[170] Respondent Document Submission [redacted].

[171] *Troubleshoot common Azure deployment errors*, MICROSOFT LEARN (June 20, 2024), https://learn.microsoft.com/en-us/azure/azure-resource-manager/troubleshooting/common-deployment-errors.

blocking" the AI developers.[172] This level of service and engineering may come at an additional cost to a typical developer not engaged in the partnerships.[173]

### 4.4.4 - The partnerships provide CSP partners access to performance and financial data of AI models and infrastructure needs of AI model developers.

The partnerships provide CSP partners with access to confidential and potentially sensitive financial and performance information.

For example, internal company documents showed that at least one CSP partner had access to first-party API [redacted] and customer counts of its AI developer partner.[174] This access to data regarding performance of the model developer's first-party API might not be accessible to the CSP absent the partnership.[175] In one case, this data [redacted]

> "[redacted] questions explicitly from our leadership. [redacted] usage metrics for [redacted] including customer [redacted]"[176]

The CSP also viewed weekly revenue trend data and customer updates.[177]

The partnerships also provide CSP partners with access to confidential and potentially sensitive financial information. This exchange of financial and performance information was explicitly required by contract in at least one case.[178]

The confidentiality of the AI developers' financial information was underscored by one study respondent, who characterized the nature of AI developers' financial information as "very sensitive."[179]

Access to this information may be helpful to CSPs as they develop their long-term strategic plans. For example, in internal documents, one CSP partner described its partnership with an AI model developer

---

[172] [redacted].

[173] *See AWS Premium Support*, AWS, https://aws.amazon.com/premiumsupport/ for an example of the type of support available from a CSP at an extra cost.

[174] *See, e.g.,* Respondent Document Submission [redacted].

[175] Similarly, for example, information exchanged regarding MaaS revenue on a rival CSP's MaaS platform would be notable because it might not be accessible to the CSP partner otherwise. This contrasts with the MaaS revenue of the AI developer on its CSP partner's own platform, which the CSP would have access to for billing purposes. Respondent Document Submission [redacted].

[176] Respondent Document Submission [redacted].

[177] Respondent Document Submission [redacted].

[178] *See, e.g.,* [redacted]

[179] Respondent Document Submission [redacted]

partner as providing a "multi-year crystal ball into the future needs of AI infrastructure"[180] and that by choosing "to partner closely with [redacted], our silicon, networks, systems software, and data center infrastructure will be significantly better-positioned in the marketplace to compete and win against tough competition from [redacted]."[181]

Strategic planning decisions that CSPs must make include whether and where to build data centers, and with what capabilities. Documents described the CSP partners gaining non-public insights into these other cloud needs of state-of-the-art AI systems. For example, [redacted] Parity and exclusivity provisions—as well as employee exchanges—may affect the impact of this and other knowledge-sharing.

### 4.4.5 - The partnerships permit CSP and AI developer partners to share or otherwise obtain training data.

For instance, at least one CSP partner had access to the outputs of its partner's models and planned to use that output to train its own models.[183] The CSP wrote internally that access to this data "could both accelerate our development efforts and also help us understand and value the benefits of using high-quality training data."[184] Another CSP partner expressed interest in accessing its AI developer partner's data directly.[185]

Similarly, AI developer partners may be able to access additional training data with the help of their CSP partner. The CSP partner respondents offer popular products in search, social media, advertising technology, productivity software, and streaming platform verticals, among others. These companies generally collect and store large amounts of raw and labelled text, audio, image, and video data for each of those products. Much of this data lies within the CSP partner's own systems, meaning the data is either not visible to the public (*e.g.*, form entries or direct message text) or is visible to the public but protected from scraping by purported legal or technical means.[186] Some of the CSP respondents in this study also operate large-scale web crawlers that serve their search engine products.[187]

[180] Respondent Document Submission [redacted].
[181] Respondent Document Submission [redacted].
[182] Respondent Document Submission [redacted].
[183] Respondent Document Submission [redacted].
[184] *Id.*
[185] Respondent Document Submission [redacted].
[186] *See, e.g.,* Michael Dinzinger et al., *A Survey of Web Content Control for Generative AI*, ARXIV E-PRINTS (Apr. 2, 2024), https://arxiv.org/pdf/2404.02309.
[187] *Googlebot*, GOOGLE (November 26,2024), https://developers.google.com/search/docs/crawling-indexing/googlebot; *Overview of Bing crawlers (user agents)*, MICROSOFT BING, https://www.bing.com/webmasters/help/which-crawlers-does-bing-use-8c184ec0.

Additionally, AI developers may obtain training data for their models by various routes, including scraping training data from the web,[188] using data collected from users of other products to train their models,[189] hiring short-term contractors to create the data,[190] buying data from data brokers or third-party companies,[191] or buying it from rightsholders directly.[192]

In at least one case, an AI model developer partner sought or planned to seek access to training data from the CSP partner's own products and services.[193] One CSP partner noted that its first-party data would be used by its AI developer partner, writing that "[redacted] customer data and traffic in training and inferencing [redacted]."[194]

Training data is a key input that can impact the development and performance of AI models. In at least one case, a CSP partner discussed developing products that improved the ability of AI developers to obtain training data.[195] For instance, [redacted][196]

### 4.4.6 - The partnerships include co-development plans for CSP-designed semiconductor chips optimized for the needs of AI developer partners' models.

The partnerships also feature co-development plans for CSP-designed AI chips, allowing for testing and feedback from AI developers to improve the chips for applications in AI training and inference. For example, Anthropic has publicly described its "deep collaboration" with AWS on chip development, writing that "Anthropic is working closely with Annapurna Labs at AWS on the development and optimization of future generations of Trainium accelerators, advancing the capabilities of specialized machine learning hardware."[197]

---

[188] Comment of the United States Federal Trade Commission, United States Copyright Office Docket No. 2023-6 (October 30, 2023), https://www.ftc.gov/system/files/ftc_gov/pdf/p241200_ftc_comment_to_copyright_office.pdf.
[189] FTC Staff in the Office of Technology and the Division of Privacy and Identity Protection, *supra* note 95.
[190] *See, e.g.,* Krystal Kauffman & Adrienne Williams, *Turk Wars: How AI Threatens the Workers Who Fuel It*, STANFORD SOCIAL INNOVATION REV. (Oct. 11, 2023), https://ssir.org/articles/entry/ai-workers-mechanical-turk.
[191] Katie Paul & Anna Tong, *Inside Big Tech's underground race to buy AI training data*, REUTERS (Apr. 5, 2024), https://www.reuters.com/technology/inside-big-techs-underground-race-buy-ai-training-data-2024-04-05/.
[192] *Partnership with Axel Springer to deepen beneficial use of AI in journalism*, OPENAI (Dec. 13, 2023), https://openai.com/index/axel-springer-partnership/.
[193] *See, e.g.,* Respondent Document Submission [redacted]
[194] Respondent Document Submission [redacted].
[195] Respondent Document Submission [redacted]
[196] *Id.*
[197] *Powering the next generation of AI development with AWS*, ANTHROPIC (Nov. 22, 2024), https://www.anthropic.com/news/anthropic-amazon-trainium.

Multiple CSPs cited this as a feature of the collaboration in both their public and private documents.[198] In at least one partnership, a participant asserted that chip feedback and co-development can benefit the CSP, estimating that their chip co-design efforts resulted in an improvement in price-performance metrics for CSP partners.[199]

Furthermore, at least one CSP partner viewed the partnership as an opportunity to ensure that a major AI model developer was not training its models solely on a competitor's chips.[200]

CSPs have been developing AI chips for their own use and to offer on their cloud platforms.[201] CSP partnerships with AI developers to build chips could potentially allow them to develop chips more quickly—and to develop chips that are tailored to the needs of the major AI developers.

AI developer partners could also potentially benefit from chip co-development with their CSP partners, such as hardware infrastructure optimized for their training and inference needs. OpenAI has publicly described its hardware and software collaboration with Microsoft as "pav[ing] the way for training more capable models and making those models cheaper for our customers."[202]

## 4.5 - The partnerships offer CSP and AI developer partners opportunities to expand current products.

### 4.5.1 - The partnerships offer CSP partners the ability to integrate their AI developer partners' AI models into their existing products and tools.

In addition to the revenue created in some cases by offering access to the AI developers' models,[203] some CSP partners emphasized the importance of ensuring that "the most capable models in the world"[204] are available on their cloud platforms. One CSP cited the expansion of its model-as-a-service ("MaaS")

---

[198] *What you need to know about the AWS AI chips powering Amazon's partnership with Anthropic*, AMAZON (October 16, 2023), https://www.aboutamazon.com/news/aws/what-you-need-to-know-about-the-aws-ai-chips-powering-amazons-partnership-with-anthropic. ("Anthropic will train and deploy their future FMs on the AWS Cloud using our Trainium and Inferentia chips. The AWS chips will be the engines behind the FMs. In addition, Anthropic will collaborate with us in the development of future Trainium and Inferentia technology."); *See also* Siegel, *supra* note 45 ("'Since first partnering with Microsoft, we've collaborated to co-design Azure's AI infrastructure at every layer for our models and unprecedented training needs,' said Sam Altman, CEO of OpenAI. 'We were excited when Microsoft first shared their designs for the Maia chip, and we've worked together to refine and test it with our models. Azure's end-to-end AI architecture, now optimized down to the silicon with Maia, paves the way for training more capable models and making those models cheaper for our customers.'"); [redacted].

[199] Respondent Document Submission [redacted].

[200] Respondent Document Submission [redacted]

[201] *See, e.g.,* Norm Jouppi, *Google supercharges machine learning tasks with TPU custom chip*, GOOGLE CLOUD (May 18, 2016), https://cloud.google.com/blog/products/ai-machine-learning/google-supercharges-machine-learning-tasks-with-custom-chip.

[202] Siegel, *supra* note 45.

[203] *See* Section 4.1.

[204] Respondent Document Submission [redacted].

offerings as a benefit of the partnerships.[205] As a reminder, MaaS platforms allow customers to access multiple AI models through a centralized platform. A press release from Amazon, for example, makes clear that the integration of Anthropic's Claude models into Amazon's Bedrock is part of the larger partnership that includes chip co-development, investments, and cloud commitments.[206]

Public reporting and documents produced by respondents indicated that some of the partnerships created avenues for CSP partners to integrate AI developers' models into their own consumer-facing products, and that some of these integrations are in progress or have been completed.[207]

For partnerships involving brand association, CSP partners viewed association with a leader in AI model development in a quickly growing sector as providing benefits. For instance, one partnership provides for a CSP to be described as the primary cloud provider to the AI partner.[208] Another CSP partner described the benefit of partnering with its AI developer company as follows: "our association with [redacted] and the [redacted] brand has attracted customers to explore [redacted] as their go-to platform for training and large language models, boosting our leadership position [redacted]."[209] The parameters of branding associations were a negotiation point in at least one partnership.[210]

### 4.5.2 - The partnerships offer AI developer partners potential paths to growth through the CSP partners' cloud platforms or product integrations with CSP products.

The partnerships generally allow the AI developer partners to offer their models via the CSP partner's cloud platforms as well as through software product integrations.

As a part of the partnerships, AI developer partners made a number of their models available on MaaS platforms owned by their CSP partner, as discussed above.[211] MaaS platforms allow customers to access multiple AI models through a centralized platform. Having models available on major cloud platforms may provide business growth opportunities through exposure to businesses that use those platforms already.

Documents obtained through the 6(b) study suggest that AI developer partners sought in part to grow their business through access to CSP platforms, including by increasing their access to enterprise customers.[212]

---

[205] Respondent Document Submission [redacted] (citing the "[o]pportunity to expand [redacted] with [AI developer partner] [redacted] models").

[206] *Amazon and Anthropic Announce Strategic Collaboration to Advance Generative AI*, AMAZON (September 25, 2023), https://press.aboutamazon.com/2023/9/amazon-and-anthropic-announce-strategic-collaboration-to-advance-generative-ai.

[207] Britney Nguyen, *Amazon's AI-powered Alexa is reportedly not even close to being ready*, QUARTZ (June 14, 2024), https://qz.com/amazon-alexa-ai-apple-openai-chatgpt-1851540465; Jared Spataro, *Introducing Microsoft 365 Copilot - your copilot for work*, MICROSOFT (Mar. 16, 2023), https://blogs.microsoft.com/blog/2023/03/16/introducing-microsoft-365-copilot-your-copilot-for-work/; *Use Copilot in Microsoft Teams meetings,* MICROSOFT, https://support.microsoft.com/en-us/office/get-started-with-copilot-in-microsoft-teams-meetings-0bf9dd3c-96f7-44e2-8bb8-790bedf066b1.

[208] *See, e.g.,* Amazon, *supra* note 206.

[209] Respondent Document Submission [redacted].

[210] *See, e.g.*, Respondent Document Submissions [redacted]

[211] *Use Anthropic's Claude models*, GOOGLE CLOUD (Nov. 26, 2024), https://cloud.google.com/vertex-ai/generative-ai/docs/partner-models/use-claude; *Anthropic's Claude in Amazon Bedrock*, AWS, https://aws.amazon.com/bedrock/claude/; Eric Boyd, *Introducing GPT-4 in Azure OpenAI Service,* MICROSOFT.

[212] [redacted]

Furthermore, in at least one case, a respondent expressed that an AI developer partner had the intention to prioritize deployment on CSP model-as-a-service platforms above deployment of the developer's own first-party services.[213]

Another way for AI developers to access customers is direct integration of AI developers' models into enterprise or consumer-facing products offered by CSPs, such as productivity software or social media tools, as described above.[214] As also noted above, company statements have mentioned some of these integrations, with one CSP partner integrating its partner's model into productivity and search tools.[215] In a separate instance, another CSP partner reportedly integrated its partner's AI model into a home voice assistant.[216]

## 5. Areas to Watch Regarding Potential Implications of the AI Partnerships

Section 4 outlined staff observations regarding the key terms of the partnerships. The partnerships might impact competition in myriad ways. The impact could be on competition among CSPs, including either existing CSPs or potential future entrants. There could also be potential impacts on competition between AI developers broadly defined, including on potential future entrants, or on products incorporating AI.

To assess the impact on competition one would need to gather case specific facts, beyond those gathered in this study, such as information from other market participants that were not respondents in this 6(b) study. The analysis of the impact of the partnerships on competition would be highly fact specific and would need to reflect a broader assessment of the market conditions related to the technologies discussed in this report. FTC staff has previously underscored the fact-specific nature of competition inquiries into contractual partnerships, stating that partnerships can be preferable to mergers in some cases, but that in others, vertical or horizontal agreements may cross the line into anti-competitive conduct.[217] Furthermore, such an analysis would also rely on established antitrust principles, such as those outlined in the 2023 Merger Guidelines. Therefore, an analysis of the impact that these partnerships might have on competition is beyond the scope of this report.

This section highlights certain aspects of the partnerships that may warrant further consideration. Specifically, the following considerations are highlighted: access to certain inputs, like computing resources and engineering talent; switching costs; and access to technical and business information.

---

[213] Respondent Document Submission [redacted] *See also* Respondent Document Submission [redacted] API team focused heavily on [redacted] and [redacted] at cost of API growth and optimization."

[214] Respondent Document Submission [redacted]

[215] Yusuf Mehdi, *Confirmed: the new Bing runs on OpenAI's GPT-4*, MICROSOFT BING (Mar. 14, 2023), https://blogs.bing.com/search/march_2023/Confirmed-the-new-Bing-runs-on-OpenAI%E2%80%99s-GPT-4; *Announcing Deploy To Teams from Azure OpenAI Studio,* MICROSOFT AZURE AI SERVICES BLOG (July 25, 2024), https://techcommunity.microsoft.com/t5/ai-azure-ai-services-blog/announcing-deploy-to-teams-from-azure-openai-studio/ba-p/4198388.

[216] Greg Bensinger, *Ask Claude: Amazon turns to Anthropic's AI for Alexa revamp*, REUTERS (Aug. 30, 2024), https://www.reuters.com/technology/artificial-intelligence/amazon-turns-anthropics-claude-alexa-ai-revamp-2024-08-30/.

[217] *See* Fed. Trade Comm'n, Dealings with Competitors, https://www.ftc.gov/advice-guidance/competition-guidance/guide-antitrust-laws/dealings-competitors; Fed Trade Comm'n, Dealings in the Supply Chain, https://www.ftc.gov/advice-guidance/competition-guidance/guide-antitrust-laws/dealings-supply-chain.

The confidentiality restrictions inherent in 6(b) studies may limit specific public analysis of each partnership.

## 5.1 - The partnerships could affect access to certain inputs, such as computing resources and engineering talent.

The partnerships might affect access to key inputs such as computing resources and talent in ways that could impact competition for both AI developer partners and non-partner AI developers.

The partnerships reviewed are notable for the fact that the investors are also suppliers of an input that is key for AI developers—computing resources. Users of that input include the CSP partners themselves in their role as AI developers, the AI developer partners, and non-partner AI developers.

As noted in Section 4, the agreements give CSP partners equity shares in AI developers as well as revenue-sharing rights and certain consultation rights to varying degrees. Therefore, the partnerships might alter incentives for CSP partners that control a product or a service that other market participants would use. For example, a CSP partner might consider limiting access to computing resources for AI developers other than its partner AI developer.

Conversely, the terms of certain partnership agreements could restrict or raise the costs of using non-partner cloud providers for AI developers who are parties to the agreement. Such an agreement might limit access of competing CSPs to these large customers thereby limiting these CSPs' ability to expand. These potential restrictions on competing non-partner CSPs may make it difficult for AI developers—both partners and non-partners—to obtain the computing resources they need to train models at scale and run inference.

Whether or not these dynamics will raise competition concerns will depend, among other things, on the availability of substitutes, the competitive significance of the product, and the competitive conditions among CSPs and AI developers. It could also depend on whether the partnerships allow the CSP and the AI developers to better coordinate their product offerings (Section 4.5).

CSP partners have acknowledged in public and private documents that AI computing resources currently have few providers (*see* Section 3.2) and are in high demand. As one CSP wrote in internal documents, "we face a problem today where scarce GPU resources are being disproportionately used by a few large customers who are getting steep discounts (up to [redacted] discounts on [redacted] is [sic] driving hoarding behavior..."[218]

CSP partners have acknowledged capacity constraints for AI computing resources publicly. On a July 2024 earnings call, Microsoft referenced "existing AI capacity constraints" and stated plainly that "we are constrained on AI capacity."[219] Amazon stated similar concerns in an October 2024 earnings call, citing demand for AI chips as its primary AI capacity concern: "Pretty much everyone today has less [AI]

[218] Respondent Document Submission [redacted].
[219] Nadella & Hood, *supra* note 74. ("We expect capital expenditures to increase on a sequential basis given our cloud and AI demand, as well as existing AI capacity constraints. […] We are, and we've talked about it now for quite a few quarters, we are constrained on AI capacity.").

capacity than they have demand for, and it's really primarily chips that are the area where companies could use more supply."[220]

This is happening in an environment where there are indications that the large computing requirements of the partnerships may have, at times, made it difficult for CSP partners to have computing resources left for other AI developers, including startups. In 2023, one CSP partner wrote internally about its inability at that time to meet AI-specific demand even from smaller AI developers, writing, "We are also unable to serve a growing number of startups who need bursts of GPU capacity to train or fine tune deep-learning models and bring their Generative AI products to market, but can't commit to long term deals."[221] It is unclear whether or how that CSP's ability to provide GPU capacity to startups has changed since then.

In late 2023, the shortage of computing resources for AI was so severe that AI developers reported multi-month waits for renting AI accelerator chips on major cloud providers.[222] One developer told reporters at that time that "'It is literally not possible to get access' to AI servers 'unless you have some existing contract with [major cloud providers] or you're pre-paying for it.'"[223]

### 5.1.2 - Access to talent

AI developer talent is also important to researching and developing the most advanced AI models and applications. As FTC staff has previously stated, "the talent companies can acquire and maintain may play a key role in not only the path, but also the rate, of generative AI's evolution."[224] An open question is whether the partnerships may consolidate access to this talent pool in the hands of a limited number of firms.

One 6(b) respondent wrote that the "talent bar" is "exceptionally high," and that a key organizational objective for 2024 is ensuring that the company has access to this rare talent.[225] In another situation, a respondent wrote in internal documents about development of large AI models that "[t]he scaling frontier is primarily about money and compute, but of course also depends strongly on talent."[226]

While specialized skills are core to many business endeavors—including software engineering—the skills necessary to develop and deploy large-scale generative AI models are relatively rare and may be difficult to acquire outside of working for large AI developers or the hyperscalers themselves.[227] Underscoring the relative scarcity of certain types of AI talent, one 6(b) respondent wrote that "For training the LLM, there

---

[220] *Amazon.com (AMZN) Q3 2024 Earnings Call Transcript*, MOTLEY FOOL (Oct. 31, 2024), https://www.fool.com/earnings/call-transcripts/2024/10/31/amazoncom-amzn-q3-2024-earnings-call-transcript/ ("I think pretty much everyone today has less capacity than they have demand for, and it's really primarily chips that are the area where companies could use more supply. And so, we're growing at a very rapid rate and have grown a pretty big business here in the AI space.").

[221] Respondent Document Submission [redacted].

[222] Holmes & Gardizy, *supra* note 70.

[223] *Id.*

[224] Staff in the Bureau of Competition & Office of Technology, Fed. Trade Comm'n, Generative AI Raises Competition Concerns (June 29, 2023), https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns.

[225] Respondent Document Submission [redacted].

[226] Respondent Document Submission [redacted].

[227] FTC Staff in the Bureau of Competition & Office of Technology, *supra* note 224. ("Developing a generative [AI] model requires a significant engineering and research workforce with particular—and relatively rare—skillsets, as well as a deep understanding of machine learning, natural language processing, and computer vision."); *see, e.g.*, this suggested skill in a recent OpenAI job posting: "Past experience in creating high-performance implementations of deep learning algorithms" here: *Research Engineer*, OPENAI, https://openai.com/careers/research-engineer/.

are only a few groups in the world who understand the methodology [redacted]."[228]

Dynamics around access to talent may be reflected in some recent deals between AI startups and CSPs, which have included elevating talent from AI startups to leadership roles at CSPs. For example, Microsoft's March 2024 deal with Inflection led to Inflection's former CEO becoming the CEO of Microsoft AI;[229] Google's August 2024 transaction with Character.ai reportedly resulted in Character.ai's former CEO becoming a VP at Google, where he is "one of three people leading Google's efforts to build the next version of its most powerful AI technology, Gemini;"[230] and Adept's former CEO became the head of one of Amazon's AI units as a part of a deal between Adept and Amazon in June of 2024.[231]

## 5.2 - The partnerships could increase contractual and technical switching costs for AI developer partners.

The partnerships might affect AI developer partners by increasing switching costs and making it more difficult for them to change CSPs or by restricting their use of multiple CSPs.

As mentioned previously, AI developers require a significant amount of computing resources and AI-specific infrastructure. If one CSP cannot meet an AI developer's needs, then an AI developer might consider contracting with one or multiple other CSPs. However, exclusivity terms and other provisions in the partnership agreements may prevent the AI developer partner from adding resources from a different CSP or result in increased switching costs that makes migrating from one CSP to another very difficult.

The partnership contracts generally include cloud computing spending commitments, and some include publicly announced exclusivity requirements that require AI developers to use services from their CSP partners.[232]

As described in Section 4.3, AI developer partners have agreed to spend a significant and set amount of investment money on services from their CSP investor, sometimes over a prescribed amount of time. For instance, according to public reporting, Anthropic's September 2023 committed use contract with Amazon requires Anthropic to spend $4 billion on AWS over the next five years.[233] These spend

[228] Respondent Document Submission [redacted]
[229] Satya Nadella, *Mustafa Suleyman, DeepMind and Inflection Co-Founder, Joins Microsoft to Lead Copilot*, MICROSOFT (2024), https://blogs.microsoft.com/blog/2024/03/19/mustafa-suleyman-deepmind-and-inflection-co-founder-joins-microsoft-to-lead-copilot/ ("Mustafa will be EVP and CEO, Microsoft AI, and joins the senior leadership team (SLT), reporting to me.").
[230] Miles Kruppa & Lauren Thomas, *Google Paid $2.7 Billion to Bring Back an AI Genius Who Quit in Frustration*, WALL STREET JOURNAL (Sept. 25, 2024), https://www.wsj.com/tech/ai/noam-shazeer-google-ai-deal-d3605697 ("It is a remarkable turn of events after Shazeer publicly said the search giant had become too risk-averse in developing AI. The 48-year-old engineer is now one of three people leading Google's efforts to build the next version of its most powerful AI technology, Gemini. […] Shazeer, however, is already back to work at Google with the title of vice president.").
[231] Palmer, *supra* note 35. ("Luan will oversee Amazon's 'AGI Autonomy' division, and report to Prasad, he wrote in the memo, which CNBC obtained. Amazon confirmed the contents of the memo. Geekwire was first to report on it.").
[232] OpenAI, *supra* note 149.
[233] *See* Jin & Dotan, *supra* note 7.

commitments may limit the options of AI developers and, for example, may require them to stay in the cloud ecosystem if they are to receive future investments from the CSP.[234]

Additionally, as described in Section 4.2, multiple partnerships have provisions that impose conditions or restrictions on AI developer partners' ability to operate with other cloud providers or companies. For example, OpenAI and Microsoft have publicly disclosed that Microsoft is OpenAI's "exclusive cloud provider."[235] These restrictions could keep AI developer partners from switching CSPs or working with multiple CSPs, even after the partnerships have ended. Exclusive provision of computing resources could have other long-term effects. For example, even if Microsoft's supercomputer, described in Section 4.4.1, were to become available for other developers in the future, it could potentially be more difficult for those developers to adapt their models to the supercomputer's design, due to the fact that it was "designed specifically to train [OpenAI's] AI models."[236]

The combination of specific contractual requirements of the partnerships described in this report could potentially have the effect of turning AI developer partners into long-term customers for CSPs.

The committed spend agreements are also made in an environment where AI developer partners may have limited options, both in terms of CSPs and funders. Although CSPs are not the only source of funding and computing resources for AI developers, they remain by far the largest for the AI developer partners in this study.[237] To the extent that the AI developer partners feel they must partner with a CSP—and to the extent they feel they have a limited number of CSPs with which to partner—any potential benefits to AI developer partners from the partnerships do not eliminate concerns arising from the potential lock-in effects of the partnership agreements.

Because AI developers are engaged in a complex set of agreements with CSPs—including billions of dollars of investments—AI developers are likely also considering the terms of the broader partnership. One consideration could be how longer term spend commitments are packaged with investments and computing resources in the partnerships, potentially resulting in AI developers accepting longer spend commitments than they might otherwise. Similarly, it is not clear whether AI developers could obtain computing resources and capital without needing to commit to other partnership terms.

Egress fees on the vast stores of data used for AI—which were raised as a potential concern by commenters in response to the FTC's 2023 Cloud Computing RFI[238]—could be another variable considered by AI developers when determining the cost of switching providers after the partnerships have ended.

---

[234] Anissa Gardizy & Kate Clark, *Anthropic Has Floated $40 Billion Valuation in Funding Talks*, THE INFORMATION (Sept. 23, 2024), https://www.theinformation.com/articles/openai-rival-anthropic-has-floated-40-billion-valuation-in-early-talks-about-new-funding.

[235] *See, e.g.*, Microsoft, *supra* note 149.

[236] Langston, *supra* note 158.

[237] Respondent Document Submission [redacted]

[238] *See* Nick Jones, Fed. Trade Comm'n, Cloud Computing RFI: What we heard and learned (Nov. 16, 2023), https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/11/cloud-computing-rfi-what-we-heard-learned.

### 5.2.2 - Technical Switching Costs

Technical barriers may also impact AI developers' ability to easily switch to alternative CSPs.[239] The time and cost of switching may be especially important in the context of generative AI due to the rapid development and deployment of new models. For instance, certain AI chip types are only available via certain CSPs—for example, Amazon and Google each design AI chips that are only available to access via their cloud offerings.[240] In these cases, if an AI developer relies on those chips, it would need to incur both cloud and chip switching costs to switch CSPs.

Regarding cloud switching costs, migration from one CSP to another can be potentially time- and capital-intensive due to a range of factors, including technical barriers. This may be particularly true for AI developers that rely on CSPs for specialized computing resources and managed services. OpenAI itself was reportedly required to "port its services to run on Microsoft Azure" as part of its initial partnership with Microsoft in July of 2019.[241] Changes made to OpenAI's publicly available code suggest that OpenAI was still in the process of migrating to Azure in December of 2020, a total of 17 months later.[242]

Regarding AI chip switching costs, as explained in Section 3, there are a number of factors that can make it difficult to switch between AI chips, including that models may be built in ways that depend on specific AI chips. For example, the software used to interact with AI chips can be specific to the chip type, meaning that AI developers would need to rewrite the software that interacts directly with AI chips if they wanted to switch AI chips. This is true of Nvidia's software platform, CUDA, which is currently only compatible with Nvidia chips;[243] and AWS's Neuron Software Development Kit, which currently only works with AWS's Inferentia and Trainium chips.[244]

Thus, migrating to infrastructure that uses other AI chip types could potentially be time-intensive and costly. As public reporting recently described with respect to Anthropic's use of Amazon's specialized AI chips, "Shifting to the Amazon server chip ... could also lock Anthropic into using Amazon Trainium servers, making it more difficult for the AI startup to use other cloud providers or to lease its own data centers in the future, as Amazon doesn't make its hardware available to facilities run by other companies."[245]

## 5.3 - The partnerships provide CSP partners access to sensitive technical and business information that may be unavailable to others.

The partnership agreements may allow CSP partners to use certain sensitive technical and business information that other CSPs and AI developers may not have (*e.g.*, generative AI models, AI development

---

[239] The AI developers subject to this report have to date only used CSPs in order to access the large scale of computing resources needed to train and deploy their models; thus, the focus of this section is on switching CSPs.
[240] Smith, *supra* note 56.
[241] *OpenAI forms exclusive computing partnership with Microsoft to build new Azure AI supercomputing technologies*, MICROSOFT (July 22, 2019), https://news.microsoft.com/2019/07/22/openai-forms-exclusive-computing-partnership-with-microsoft-to-build-new-azure-ai-supercomputing-technologies/. ("OpenAI will port its services to run on Microsoft Azure, which it will use to create new AI technologies and deliver on the promise of artificial general intelligence[.]").
[242] Jeffrey Wu, *Commit to openai / gpt-2 repository*, GITHUB (Dec. 2, 2020), https://github.com/openai/gpt-2/commit/a74da5d99abaaba920de8131d64da2862a8f213b. (The change made on December 2, 2020 includes "move to azure" as the commit message, a message added in version control systems to explain why a change has been made to code.).
[243] *CUDA GPUs – Compute Capability,* NVIDIA, https://developer.nvidia.com/cuda-gpus.
[244] *Welcome to AWS Neuron*, AWS, https://awsdocs-neuron.readthedocs-hosted.com/en/latest/index.html. ("AWS Neuron is the SDK used to run deep learning workloads on AWS Inferentia and AWS Trainium based instances.").
[245] Gardizy, *supra* note 126.

methods, confidential chip co-design, partner finances, and customer usage and revenue numbers) to develop their own internal models and applications, including products that may compete with those of their partners.

As noted in Section 4, CSPs, through their partnerships with AI developers, gain access to business intelligence and product information that they might not otherwise have. This includes information that AI developer partners are contractually required to provide to CSP partners or information that the CSPs may gain via monitoring, analyzing, and collecting information from their AI customers (*e.g.*, usage metrics).

In addition to offering cloud services, all three CSP partners develop their own generative AI models and products that may compete with those of their partners. One CSP partner noted that its AI developer partner "continues to feel competitive threat from" the CSP.[246]

Voice assistants are one example—of many—of an area of potential overlap between CSP partner products and AI developer partner products. One reporter in 2023 described a risk that new voice assistants, such as those provided by AI developer partners, might be "coming for Siri and Alexa's jobs."[247] In fact, Siri and Alexa are now, or will soon be, reportedly powered by the AI partner respondents in this study.[248] Potential concerns about information exchange between CSPs and their AI model developer partners may be heightened by any such competition between their products.

Technical knowledge from the AI developer partner—including IP and know-how such as model development methodologies and chip design suggestions—may therefore present CSP partners with valuable information regarding generative AI development unavailable to non-partner CSPs. As described in Section 4, the partnerships involved the sharing of:

a) Generative AI models;
b) AI model development methodologies, for instance via explicit model sharing agreements and employee exchanges;
c) Confidential chip co-design; and
d) Miscellaneous product improvement suggestions in areas such as networking, storage, and MaaS offerings.

These information exchanges occurred through regular communication between partner companies, including emails and inter-team meetings.

Cross-company meetings and employee exchange potentially allow for exchange of business strategy information. As one company wrote in narrative responses to the agency, "To facilitate any of [redacted]'s engagements with [redacted], [redacted] may also meet with corresponding product and business team members from [redacted] to address questions or considerations; to provide ideas, suggestions, and feedback on implementation; or

[246] Respondent Document Submission [redacted]

[247] Saritha Rai, *Smarter AI Assistants Are Coming for Siri and Alexa's Jobs*, BLOOMBERG (May 9, 2023), https://www.bloomberg.com/news/newsletters/2023-05-09/siri-alexa-google-assistant-threatened-by-smarter-ai-chat-bots.

[248] *OpenAI and Apple announce partnership to integrate ChatGPT into Apple experiences*, OPENAI (June 10, 2024), https://openai.com/index/openai-and-apple-announce-partnership/; Jess Weatherbed, *Amazon's new Alexa voice assistant will use Claude AI*, THE VERGE (Aug. 30, 2024), https://www.theverge.com/2024/8/30/24232123/amazon-new-alexa-voice-assistant-claude-ai-model.

to provide updates on [redacted]'s performance and financial condition."[249] Documents reviewed by staff demonstrated that potentially sensitive financial and business information regularly flowed to at least one CSP.[250] CSPs could potentially use this business intelligence information—sensitive financial and strategy information—to understand, for example, which of their partners' offerings are most in demand. They could then adjust their own product strategy accordingly.

Therefore, technical knowledge from the AI developer partner—including IP and know-how such as model development methodologies and chip design suggestions—could present CSP partners with informational advantages regarding generative AI development over non-partnered CSPs. Access to insights and information is suggested by one CSP internal strategy document that described that a key motivation for a partnership was to "[m]aintain relationship with one of the leading [redacted] AI startups."[251]

Losing that relationship, for CSP partners, could have meant falling behind on generative AI development. Or, as one CSP partner put it in internal documents: "If we wait for our own models to mature, we risk not participating in developing the necessary IP to build, operate, and secure these applications. By partnering with [redacted], we can more quickly learn the 'art' of refining and reinforcing these types of models and better understand how to effectively integrate them into applications."[252] In another internal document, a respondent described the risk of CSP partners "cloning" the research efforts of their partners.[253]

Internal board meeting notes from one CSP partner highlighted that it received early access to hardware innovations through its partnership. The respondent wrote that [redacted]"[254] Through this collaboration, the CSP was able to obtain hardware-related business intelligence.

The CSP partners may also gain additional valuable information via their role as platform providers. As described in Section 3, all three CSPs offer "Model-as-a-Service" products, where customers can access a range of models via the CSP's platform. The MaaS platforms operate as an intermediary between a developer who wants to integrate AI into its application, and a model provider who wants to monetize its model. Of note, each CSP also offers its own first-party models on its MaaS platforms—Amazon's Titan series of models,[255] Microsoft's Phi series of models,[256] and Google's Gemini series.[257]

In their role as a platform for other models (including those of their partners), the CSPs gain valuable insight into several metrics that are otherwise unavailable to rival AI developers that compete on the platform. First, they have insight into the popularity and prevalence of generative AI adoption across

---

[249] [redacted]

[250] *See, e.g.,* Respondent Document Submission [redacted] (describing revenue forecasts of partner company to board); [redacted]

[251] Respondent Document Submission [redacted].

[252] Respondent Document Submission [redacted].

[253] Respondent Document Submission [redacted].

[254] Respondent Document Submission [redacted]

[255] *Amazon Titan in Amazon Bedrock*, AWS, https://aws.amazon.com/bedrock/titan/; Amazon Staff, *Introducing Amazon* Nova, our new generation of foundation models, AMAZON (Dec. 3, 2024), https://www.aboutamazon.com/news/aws/amazon-nova-artificial-intelligence-bedrock-aws.

[256] *Phi open models*, MICROSOFT, https://azure.microsoft.com/en-us/products/phi.

[257] *Innovate faster with enterprise-ready AI, enhanced by Gemini models*, GOOGLE CLOUD, https://cloud.google.com/vertex-ai?hl=en#build-with-gemini.

industries. For example, Amazon touted in an August 2024 earnings call that "tens of thousands of companies" are using Bedrock.[258] Usage patterns and volumes from platforms give additional insight relative to those who do not operate such a platform.

Second, CSPs have information on which models are being used by their cloud customers. For example, documents received pursuant to the 6(b) orders indicate that at least one CSP has data on how often its own first party models are being used relative to third party models it offers and uses that information to make business decisions.[259]

Information on usage of generative AI models of various sizes, costs, and latency by a range of customers could potentially help inform business strategies for the CSPs—including their first-party model strategy, potentially providing them with an advantage over other AI model developers (partners or otherwise). One open question is the extent to which CSPs are seeing that certain types of models are gaining more traction and shifting their strategy to focus on developing their own similar first party models—or the extent to which they are deciding not to develop models for a given use case if they see other models being highly successful at that use case.

Notably, a CSP may have access to this potentially sensitive usage data for any AI developers offering their models on the CSP's MaaS offering—not just for those AI developers in the partnerships considered in this study. For example, Microsoft reported in May of 2024 having "nearly 1,700 models in the Azure AI catalog" to choose from.[260] The hosting of these models could potentially also give CSPs insight across non-partnership AI developer usage.

The extent and significance of these potential information advantages over non-partnered CSPs may require further evaluation, such as by evaluating what information is available to non-partnered CSPs.

# 6. Conclusion

Over the past few decades, technological developments have provided enormous benefits to society and helped the United States stay a global leader in innovation. Commission leadership has highlighted the great promise of AI to benefit consumers and make American businesses more effective,[261] while recognizing potential risks to competition and consumers, such as "the expanding adoption of A.I. […] further locking in the market dominance of large incumbent technology firms."[262]

---

[258] *Amazon.com (AMZN) Q2 2024 Earnings Call Transcript*, MOTLEY FOOL (Aug 1, 2024), https://www.fool.com/earnings/call-transcripts/2024/08/01/amazoncom-amzn-q2-2024-earnings-call-transcript/.

[259] Respondent Document Submission [redacted] Table 2 "Revenue by Service" (showing a breakdown of revenue from [redacted] vs. [redacted]).

[260] Ryan Roslansky, *Jefferies Software Conference*, MICROSOFT (June 6, 2024), https://www.microsoft.com/en-us/investor/events/fy-2024/jefferies_software_conference. ("And so we have nearly 1,700 models in the Azure AI model catalog. So that's what lives within that Azure AI Studio I mentioned earlier.").

[261] *See, e.g.*, Prepared Statement of Melissa Holyoak, Commissioner, Federal Trade Commission Before the Subcommittee on Innovation, Data, and Commerce of the Energy and Commerce Committee, United States House of Representatives, Concerning "The Fiscal Year 2025 Federal Trade Commission Budget" (July 9, 2024), https://www.ftc.gov/system/files/ftc_gov/pdf/commissioner-holyoak-testimony-7-5-24.pdf.

[262] Lina M. Khan, *Lina Khan: We Must Regulate A.I. Here's How.*, NEW YORK TIMES (May 3, 2023), https://www.nytimes.com/2023/05/03/opinion/ai-lina-khan-ftc-technology.html. ("A handful of powerful businesses control the necessary raw materials that start-ups and other companies rely on to develop and deploy A.I. tools. This includes cloud services and computing power, as well as vast stores of data. "Enforcers and regulators must be vigilant. Dominant firms could use their control over these key inputs to exclude or discriminate against downstream rivals, picking winners and losers in ways that further entrench their dominance.").

As staff has laid out above, generative AI requires extensive resources and funding to provide access to infrastructure, access to talent, and access to data to train foundation models. This is why some firms have used technology partnerships, investments, and other deals to accelerate development and deployment of AI models. While these types of deals are not a new phenomenon, they have seen a high-profile resurgence due to the generative AI wave and increased merger scrutiny.[263] The partnerships examined in the present report are prime examples of this resurgence—making it important that the FTC and the public understand the terms of the partnerships and their impacts on competition and consumers.

Through a scoped and targeted 6(b) process, staff were able to obtain core non-public pieces of information regarding partnership terms and establish areas to watch going forward related to potential impacts of the partnerships on the AI ecosystem.

[263] *See, e.g.,* Press Release, Fed. Trade Comm'n, FTC and DOJ Seek Info on Serial Acquisitions, Roll-Up Strategies Across U.S. Economy (May 23, 2024), https://www.ftc.gov/news-events/news/press-releases/2024/05/ftc-doj-seek-info-serial-acquisitions-roll-strategies-across-us-economy.

# Appendix A. Order Specifications

Complete orders, including definitions, can be found on the Commission's website.[264]

**Specifications**

1. Provide all agreements and related documents (*e.g.*, term sheet, schedules, attachments, side letters) for the Company's investment in or partnership with [PARTNER COMPANY] (the "transaction"). Identify and explain each strategic rationale for the transaction discussed or contemplated by the Company, or any of its officers, directors, or employees.

2. Identify, explain, and provide documents sufficient to show: (i) the nature and extent of Your interaction with and influence over [PARTNER COMPANY]; (ii) the engagement and interaction between You and [PARTNER COMPANY] as it concerns business operations and strategic decision-making (including, but not limited to, decisions around the initiation, launch, road-mapping, development, and discontinuation of products and services; decisions around the pricing of products and services; decisions around the granting of access to products and services; and decisions around personnel, including the hiring and allocation of staff; (iii) the subject, personnel, and dates of any regular meetings between You and [PARTNER COMPANY]; and (iv) any governance or oversight rights or options (*e.g.*, Board seats) that You have or may exercise with respect to [PARTNER COMPANY].

3. Provide all analyses, reports, studies, and surveys prepared by or for any senior management for the purpose of evaluating or analyzing the transaction with respect to its impact or potential impact (including but not limited to the impact or potential impact on competition; on other market participants; on Your market position across lines of business; and on Your plans with regards to investment, expansion, retrenchment, or allocation of resources across markets and lines of business). For avoidance of doubt, this includes all analyses, reports, studies, and surveys by You or [PARTNER COMPANY] or prepared by a third party on Your or [PARTNER COMPANY'S] behalf.

4. Provide, from January 1, 2022 to the present, all documents prepared by or for any senior management related to: (i) any agreement or expectation of exclusivity between You and [PARTNER COMPANY] across any line of business; (ii) any agreement or expectation of Your Company receiving privileged access to products or services provided by [PARTNER COMPANY] or of [PARTNER COMPANY] receiving privileged access to products or services provided by Your Company; (iii) any restrictions or conditions placed on [PARTNER COMPANY'S] ability to set pricing, set terms, or grant access to its products or services, or to make any other decision regarding business operations or strategy; (iv) competitive conditions for any cloud computing or generative AI products or services (including, but not limited to, cloud computing infrastructure, microchips, training data, models, and products or services that incorporate generative AI); (v) the use of cloud computing in generative AI products or services; and (vi) your willingness or ability to provide, supply, or offer cloud computing services or generative AI products, services, or capabilities to other companies (including, but not limited, to

[264] Fed. Trade Comm'n, *supra* note 3.

any discussions of pricing and service levels, supply limitations, technological constraints, or contractual restrictions).

5. Provide all materials, including any communications, responses, analyses, reports, studies, surveys, and any other documents, that You have provided to any government entity, including foreign government entities, in connection with any investigation, request for information or other inquiry related to Specifications 1-4, above.