Antimonopoly Tools for Regulating Artificial Intelligence

Ganesh Sitaraman and Tejas Narechania

Vanderbilt Policy Accelerator for Political Economy & Regulation

Introduction[1]

Since OpenAI released ChatGPT in November 2022, debates over regulating artificial intelligence (AI) among policymakers, technologists, and scholars have intensified. For all the interest in regulating AI, however, there has been comparatively little discussion of AI's industrial organization and market structure.[2] This is surprising because critical layers in the AI technology stack are already highly concentrated.[3] As in other areas, monopoly and oligopoly in these industries can not only distort markets, chill investment, and hamper innovation, but also facilitate downstream harms to users, threaten national security and resilience, and help accumulate private power in relatively few hands.

In this Policy Brief, which is drawn from a more extensive research paper, we argue that policymakers should use antimonopoly tools to regulate the harms that come from concentration in the AI "technology stack." These include public utility tools, including structural separations, nondiscrimination requirements, and interoperability rules; public options for cloud computing and data resources; and consideration of competition when making industrial policy investments or engaging in procurement.

Understanding the AI Technology Stack

Understanding the AI technology stack is the foundation for identifying problems that arise from market power and concentration. Drawing on accounts from industry investors and analysts, AI's technology stack can be described in four basic layers: hardware, cloud computing, models, and applications.

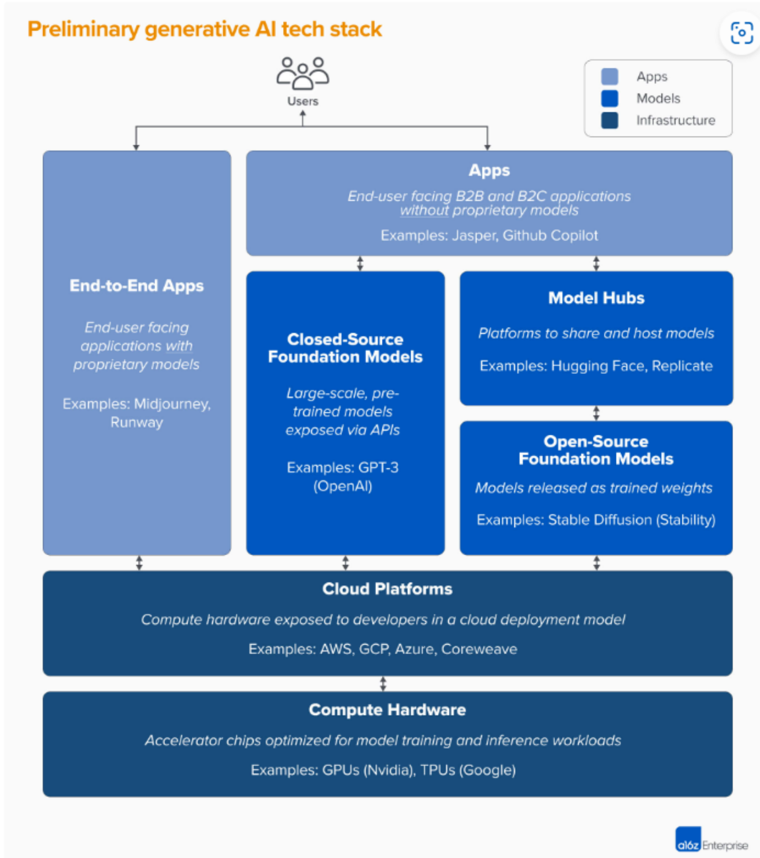

Figure 1. The Generative AI "Tech Stack" (source: Andreesen Horowitz[4])

- **Hardware.** The hardware layer includes the production of microchips and processors—the horsepower behind AI's computations. This layer is extremely concentrated, with a few firms dominating important aspects of production. This is partly because as chip technologies have become more and more sophisticated, fewer firms are able to supply the needed technologies. Nvidia, which designs chips, has captured between 80 and 95 percent market share of the GPU chip business used for AI.[5] Nvidia's chips are, in turn, manufactured (i.e., fabricated) by Taiwan Semiconductor Manufacturing Corporation (TSMC),[6] which is far and away the dominant semiconductor manufacturer.[7] Only Samsung also fabricates the smallest, highest powered chips.[8] To make the smallest chips requires photolithography equipment, something only one company in the world, the Dutch firm ASML,[9] provides—and sells for between $150 and $200 million per machine.[10]

- **Cloud Computing.** The cloud computing layer consists of the computational infrastructure that is required to host the data, models, and applications that comprise AI's algorithmic outputs. This layer, too, is highly concentrated, with three firms (Amazon Web Services (AWS), Google Cloud Platform, and Microsoft Azure) dominating the market. It features several dynamics that tend toward concentration and make sustaining competition difficult, including extremely high capital costs and significant switching costs to move from one provider to another.

- **Models.** The model layer includes data, stored in unstructured "data lakes" or more structured "data warehouses"; algorithmic models, which many think of as "AI"; and modes of accessing these models, including model hubs (where developers can download and use publicly available models) and application programming interfaces (or APIs, which allow developers to communicate with proprietary models that may not be publicly available). Some firms are fully integrated and offer all three products, which are then used to develop proprietary applications; others only offer models; and still others are more disaggregated, offering raw data or serving only as a model hub.

- **Applications.** Applications are the part of the sector that the public interacts with most directly: When we enter a prompt into ChatGPT, for example, we use an application (ChatGPT). The application draws on all prior layers in the stack: it interacts with a model (GPT4); that model is stored in a cloud computing platform (Microsoft's Azure); and that platform requires microprocessing hardware (designed by Nvidia and fabricated by TSMC). While some firms in the application layer build their products on open-source—that is, free and publicly available—models, many others offer applications built upon existing proprietary foundation models. Yet others are vertically integrated, offering both the foundational model and applications built on them. Critically, though both types of firms compete in the applications market, those that are not vertically integrated depend upon the firms that are for access to their models.

The Problems with an AI Oligopoly

Understanding the AI technology stack shows that significant portions of AI's industrial organization and market structure are likely to be, and already are, dominated by a small number of firms—and that these dominant firms are vertically integrating across the stack. This concentration—an AI oligopoly—is concerning for a variety of reasons, including abuses of power, national security and resilience risks, exacerbated economic inequality, and its effects on democracy.

Abuses of Power. Concentration across the AI stack creates opportunities and incentives for dominant firms to abuse their power, with consequences for competitors, would-be entrants, and the public. These abuses include, but are not limited to:

- *High Prices.* In hardware, dominant firms could demand monopoly and/or oligopoly prices for photolithography equipment, chip design, and chip manufacturing. Cloud computing firms might also charge monopoly or oligopoly prices. In the model layer, the high costs of obtaining good data and sufficient compute infrastructure constitute a steep barrier to entry, and foundation model providers might therefore be able to raise prices to downstream application developers for model access.

- *Self-Preferencing and Discriminatory Prices and Terms.* Monopoly or oligopoly firms at each layer in the stack may discriminate between downstream firms, offering better terms or prices to their vertically-integrated businesses as opposed to competitors. Reports already indicate that Nvidia preferences some customers over others[11] and that TSMC prioritizes its relationship with Apple over other customers.[12] This poses substantial downstream risks to competition and innovation. Model providers may also favor selected AI-based applications, including vertically integrated applications, through selectively exposed APIs.[13] In extreme cases, dominant firms could refuse to deal entirely with a customer who is also a competitor.[14] For example, Microsoft Azure could preference OpenAI, in which it has a financial stake, over other competitors, when offering access to its cloud computing infrastructure. Model providers might also favor their own applications over others by charging higher rates to third-party developers than their own in-house business lines, exclude some third-party applications from use of the model altogether, or they might prefer or advantage their own separate business lines. For example, if people ask Microsoft Bing what they should do this weekend, it might suggest playing the videogame Call of Duty—which Microsoft also owns.[15]

- *Lock-In Effects.* Cloud providers have taken steps to entrench their dominance by facilitating lock-in effects that raise the costs for consumers to switch providers through egress fees and multi-year contracts.[16] These effects exacerbate the already-high switching costs in compute, due to factors like personnel expertise in a particular platform.[17]

- *Copying*. In cloud, multiple firms have already complained that AWS has copied their product and offered their own integrated version of the product, harming their company's value and future business.[18] Firms that copy applications from competitors and incorporate them into their own offerings prevent competition and chill innovation. Venture capitalists describe this practice as creating a "kill zone," wherein the likelihood of copying or acquisition by a dominant firm discourages investment in innovative companies and products.[19]

National Security and Resilience. Concentration at critical points in the AI technology stack raises significant concerns for national security and resilience. With very few chip companies, the possibility that one foundry could be shut down due to a pandemic, weather event, war, or other emergency is significant.[20] Such an event would cause supply chain challenges for both military and non-military critical infrastructure. Concentration in cloud computing raises risks too: An oligopoly of cloud providers, integrated up and down the AI stack and without interoperability between them, gives rise to substantial software supply-chain concerns. If a cloud provider is attacked in a cyberattack, or if a cloud provider's warehouse is affected by a severe weather event, or even if an employee makes a simple mistake, dozens of AI applications—and the operations, services, and websites that depend on them—could shut down for hours, days, or longer.[21] Just as concerning is that faulty foundation models, if offered by only one firm, can lead to widespread error that could be catastrophic in emergency situations.

Economic Inequality. Like concentrated power in other industries, an AI oligopoly is likely to further economic inequality. Concentration means that a small number of firms will capture the vast majority of the sector's profits. And while it is too early to tell exactly how the introduction of AI at scale will change labor markets, it could very well create a "bifurcated job market that squeezes out the middle class."[22]

Democracy. An AI oligopoly can also contribute to democratic erosion. Concentration in AI may give a relatively small number of companies an outsized influence over the information ecosystem, complementing their outsized political influence through lobbying and other forms of political influence. Economic power also often translates into political power, as demonstrated by a voluminous literature in political science showing the extensive influence of the wealthy and interest groups on American politics.[23]

Antimonopoly Public Utility Tools

Aspects of the AI sector share features with public utilities: they are essential inputs into downstream activities, are means to an end rather than ends in themselves, and feature high capital costs, network effects, and economies of scale. Indeed, machine learning itself, as one of us has shown elsewhere, has the characteristics of a natural monopoly, even under narrow definitions.[24]

Regulatory tools from the law of networks, platforms, and utilities (NPU) have long been applied to industries that feature monopoly or oligopoly characteristics.[25] These tools can be helpful for scaling enterprises, ensuring continuity of service, preventing monopoly and oligopoly abuses, avoiding destructive competition, ensuring widespread access, promoting commercial development, and sustaining democracy.[26] We discuss here three of the most important tools from utilities law: structural separations, nondiscrimination requirements, and interoperability rules.

These regulations operate primarily *ex ante*, that is, by structuring the market to prevent harms *before* they arise. They therefore differ from traditional antitrust enforcement, which requires that harms to take place and then for litigants to bring lawsuits on a case-by-case basis to challenge those harms. In structuring a more competitive and less abusive market, public utility tools ensure that an industry develops in a way that helps foster innovation and competition.

1. **Structural Separations**. Structural separations "limit the lines of business in which a firm can engage."[27] The central benefit of structural separations is that they prevent a business from self-preferencing or leveraging their power from one business-line into another. In addition to preventing conflicts of interest and leveraging profits to gain competitive advantage in another line of business, structural separations also limit the concentration of economic power and promote a diverse business ecosystem of users of the platform.[28] Perhaps most importantly, they can be more administrable than other policies, such as nondiscrimination rules (discussed in the next section). If a company is involved in the prohibited business line, it violates the rule. This is a far simpler approach than one that permits commingling and requires monitoring behavior.

   With respect to AI, there are number of places where structural separations could be useful.[29] Perhaps most notably, structurally separating the cloud layer from higher layers in the stack could address a wide range of market dominance problems identified above. It would treat cloud computing platforms as utility providers of a service (namely, computational capacity) that is open for all kinds of uses and ensure that those providers cannot prioritize their own downstream business lines over their competitors'. Separation would likely also spur cloud providers to innovate on their cloud offerings, rather than on innovating through vertical integration.[30] This would, in turn, also facilitate innovation in downstream markets where cloud users could develop a range of products and services.

2. **Nondiscrimination, Open Access, and Rate Requirements** — One alternative to structural separation requirements are nondiscrimination and equal access rules, sometimes coupled with rate regulation.[31] Nondiscrimination rules allow a firm to operate two or more vertically-linked business lines, but require the firm to treat downstream businesses neutrally—including its own vertically-integrated business lines.[32] Equal pricing rules are an essential corollary to nondiscriminatory rules because firms could charge prohibitive prices as a workaround to evade their open access obligation.[33] In some areas, regulators have also directly set the prices firms can charge. Rate setting "is usually directed toward preventing NPU enterprises from lowering output and raising prices," while simultaneously ensuring firms earn a reasonable return on invested capital.[34]

   In the AI context, nondiscrimination and equal access rules could be adopted at multiple places in the stack. At the hardware level, given the scarcity of chips, fabricators and designers could be required to serve customers equally—at least until chip fabrication becomes more widely available. At the cloud level, cloud providers should be required to treat all downstream businesses in a nondiscriminatory fashion, be open to all comers, and offer transparent, uniform, publicly available prices. Open source and non-open source, but commercially available, data warehouses and lakes could also be subject to nondiscrimination and equal access rules. This would enable many model developers to use the data to develop and train new models. Foundation models and APIs could also be subject to such rules, so that app developers have reliable access to these resources.

3. **Interoperability Rules** — Interoperability rules lower barriers to entry and thus stimulate competition by "allowing new competitors to share in existing investments" and "imposing sharing requirements on market participants."[35] In the telecommunications context, for example, rules that required a telephone provider to transfer a user's phone number to a competing provider (and thus

required that the providers work together on an interoperable number portability system) facilitate competition by reducing switching costs for users. Those rules targeted a notable lock-in effect: It is quite cumbersome to let all your contacts know you have a new phone number.

Such requirements could be applied to parts of the AI stack, too. Policymakers might consider rules that improve interoperability among cloud platforms, easing transitions from one provider's system to another. As different providers of cloud computing services specialize—moving away from offering a pure commodity "compute" resource to more bespoke computing resources and incorporating specialized applications (or utilizing specialized hardware)—some applications developers have found it difficult to take advantage of specializations across different providers. A developer might wish, for example, to train a model on one cloud provider—but use a different one for inferential applications. Interoperability could facilitate that, potentially yielding better outcomes for participants in the downstream model and application layers, and ultimately for consumers.

Another type of interoperability rule is to mandate data sharing through federated learning. Federated learning is a technical "approach to machine learning where a shared global model is trained across many participating clients that keep their training data locally."[36] Rules that require a federated learning approach among competitors may be attractive to policymakers seeking to induce competition while ensuring that no one application, vertically integrated with the underlying model, uniquely benefits from improvements made through continuous or reinforcement learning.[37] Instead, the model's improvements are derived from all the applications that use it—and are shared among all of them, too. Such forms of AI development may help to undermine the consolidation-driving network effects of the data sublayer.

4. Entry Restrictions and Licensing Requirements – Congress has often established entry restrictions and licensing requirements for firms or individuals operating in many sectors of the economy. Such rules limit entry into a sector to firms that have registered with an appropriate regulator or otherwise have approval from the government (often in the form of a license or certificate).[38] These provisions are usually justified on one (or more) of three different grounds. First, entry restrictions can ensure safety and reliability. By placing conditions on entry into a sector, regulators can ensure that firms will operate safely and effectively. Airline pilots (and airlines themselves), for example, must be licensed. Likewise, nuclear power plants are licensed, in part, to ensure safe operation. Second, in some markets (particularly those typically characterized as natural monopolies or oligopolies) competition can lead to waste and ultimately deter capital investment.[39] In the railroad industry, for example, firms competed vigorously to build railroad tracks at a high cost—but fierce competition over price sent them into bankruptcy or merger. The result was wasted expense, abandoned rail lines, and eventual consolidation. Entry restriction can prevent these downsides, creating a stable environment for capital investment. And, third, in sectors where universal service—i.e., ensuring that everyone can access the regulated service—is a critical policy goal, regulators will often limit entry to the market.[40] This is because open competition often undermines universal service policy goals. Some services, like energy provision, have costs that vary across geographies: urban centers are typically cheaper to serve (and hence are more profitable), while rural areas can be more expensive. Without entry restrictions and related regulations, providers will tend to compete to serve the cheaper and more profitable customers (with those customers enjoying the benefits of competition), while neglecting the more expensive customer base. But entry restrictions coupled with duty-to-serve rules can ensure that everyone has access to the regulated service, often at regulated rates (typically regulated by, in part, averaging the high-cost customers with the low-cost ones).

Such requirements might be applied to the AI technology stack at various layers. First, entry restrictions might be deployed to ensure that certain foundation models and their associated applications are effective, and do not pose substantial risks to health and safety, or of bias. Indeed, the FDA's process for approving medical systems that incorporate AI resembles this approach. Similarly, licensing rules could oblige cloud providers to "know their customers," as in banking law, and ensure that entities in the model layer have checks in place to ensure non-discriminatory access, fair pricing, and safety. Applications could also be required to register with the model or cloud they use, to make it easier to identify and address dangerous or problematic behavior on a post hoc basis. Likewise, entry restrictions might help to address concerns about costly and wasteful investment—and the tendencies towards consolidation—in the model layer, which are characterized by high fixed costs, scale economies, and network effects.

A Public Option for AI

Public options are publicly-provided goods or services that coexist with private market options and are offered at a set price.[41] While the term may be most familiar from debates over health care policy, public options are common in the United States. Public schools coexist with private schools. Public swimming pools with private ones. We have parcel post from the Postal Service, public golf courses, public libraries—the list goes on.

Public options come with a range of benefits. First, they can help ensure competition, as the public option disciplines private monopolists or oligopolists that might increase prices or reduce service quality.[42] Competition from private businesses, in return, ensures that the public option provides high quality service as well.[43] Relatedly, a public option adds to the number of providers of a good or service thereby diversifying and strengthening supply chain resilience and reliability. Second, public options expand access to goods or services that might be unaffordable or scarce in the private market. A public grocery store in a rural food desert can expand access to food options in a place with few choices. A public swimming pool or playground offers a convenient place of kids to play without residents needing to buy an expensive pool or swing-set for their backyard.

As we have seen, in the AI context both of these features are important. The AI technology stack is already dominated by a small number of big technology companies, meaning that competition is limited at multiple layers. Moreover, the high cost of and high demand for semiconductors and cloud infrastructure has led to scarcity at both of these layers—meaning that some kind of prioritization is likely. Firms might prioritize vertically integrated businesses and the most profitable activities, over competitors or uses that serve the public.

1. **A Public Option for Cloud Infrastructure.** A public option for cloud infrastructure could serve as a helpful complement or alternative to tools from networks, platforms, and utilities (NPU) law—including structural separations or nondiscrimination requirements, and interoperability rules.[44] Because of high capital costs, network effects, and concerns from vertical integration, a public option for cloud could provide the cloud services that developers and end-users need—but without relying on oligopoly providers AWS, Microsoft Azure, and Google Cloud Platform. The public option could also ensure that cloud infrastructure is available at an affordable price to researchers and other users who might have different goals than private firms. A public option is also not

unworkable: the United States has a long history of publicly-run supercomputers, and Japan is in the process of building a public option supercomputer, which will make cloud services available to companies focusing on AI.[45]

While there are some existing proposals for public access to AI resources, it is unclear if these proposals are true public options or whether they will further entrench oligopoly firms at different layers in the AI stack. For example, the National Science Foundation's proposal to offer a National AI Research Resource (NAIRR) seeks to "democratize access to AI resources" and therefore "must primarily be sustained through Federal investment."[46] However, the NSF's proposal is unclear on whether NAIRR will be a public option or will simply contract with big technology companies for AI services. It suggests that NAIRR provide a mix of computational resources, including "commercial cloud" as an option.[47] It also recommends that NAIRR "include at least one large-scale machine-learning supercomputer," but then is unclear whether this would be a publicly-run resource.[48] Recently-introduced legislation to create a NAIRR states that the entity would offer "a mix of computational resources," including "on-premises, cloud-based, hybrid, and emergent resources," "public cloud providers providing access to popular computational and storage services," open source software, and APIs.[49] This structure appears to require some amount of non-oligopoly cloud provision, as the provision for an on-premises, cloud-based system is separate from the one that describes public cloud providers. But the draft legislation could be interpreted to lead only to contracts with existing cloud providers. Such a contract would further entrench their oligopoly, rather than increasing competition. It might also place public access at risk, if prices increase or if the cloud service deprioritizes public uses. The NAIRR, if created and funded, should ensure there is a true public option, rather than merely a government contract for researchers to purchase compute and other resources from the biggest cloud providers.

2. **Public Data Resources.** Data is also foundational for AI applications, and it is a resource that depends on extraordinary scale. Without considerable data—which must also be cleaned and processed—machine learning is not possible. If leading data sources are all proprietary, then the companies that control them could raise prices on downstream businesses or researchers who rely on that data for their models or applications or even deny them access, perhaps if they seek to develop a competitive product.

A public option for data could therefore "provide a pathway for start-ups and public-sector organizations to develop abilities and products."[50] A public data resource could work in a few different ways. First, the government could ensure that when it makes public data available (e.g. weather data) that the data is not privatized. Public data would thus remain in public hands—and the government would develop public data warehouses that researchers and qualified businesses could use. Second, the government could develop a data resource akin to those that private companies have developed and that would compete with it. This would likely be difficult and costly, but it would offer another avenue for competition and access. Notably, the proposed NAIRR legislation also includes provisions for data access,[51] and the federal government already has several other initiatives under consideration that are aimed at releasing public datasets to support model development.[52]

With public options at these layers, technologists would be able to develop their own foundation models or applications, without relying on the AI oligopoly for cloud services or the underlying data. This would both improve competition with those firms and ensure that public-spirited uses could be pursued.

Industrial Policy, Procurement, and Competition

Policymakers should also consider competition policy when engaging in industrial policy spending and in their procurement decisions. One of the central questions for industrial policy in the AI sector is whether investment decisions will entrench dominant players or facilitate competition. Subsidies, loan guarantees, tax advantages, and procurement decisions directed toward dominant players may keep them in positions of leadership. In areas that have a tendency toward consolidation—due to economies of scale, network effects, high capital costs, and other factors—such policies could further extend their lead. But industrial policies could also be targeted at new, smaller, and innovative actors, in which case they can facilitate competition, rather than entrench market power.[53]

1. **Industrial Policy and Semiconductors.** In the hardware layer, scarcity and supply chain vulnerability are paramount concerns. To address these problems, the United States has already taken steps to incentivize the development of chip manufacturing within the United States. The bipartisan Chips and Science Act of 2022[54] established a range of incentives to spur domestic production of cutting-edge chips. The Act committed $52.7 billion to the Departments of Commerce and Defense and the National Science Foundation to support U.S. development

of semiconductor programs.[55] The Commerce Department's Chips for America program seeks to use federal funds to crowd-in private sector investment in order to develop at least two large scale clusters for fabrication of chips.[56]

As the Commerce Department develops its program, it should carefully assess whether federal funding will entrench power or increase competition. Government officials coordinating industrial policy efforts should consider market diversification and competition as a critical element in evaluating candidates for federal grants.[57]

2. **Procurement decisions.** Federal departments and agencies are also likely to make procurement decisions with private companies and consulting firms. These decisions could range from outsourcing development of AI applications to contracts with infrastructure providers. Here too, federal officials should consider the extent to which they can promote competition, rather than entrench dominance, when making these contracts. In particular, they may have flexibility to draft contracts in a way that allows them to build in pro-competition provisions. As just one example, departments and agencies could, where possible, require that public data is quarantined from privately held data, so that the dominant platforms do not leverage their public data resources to support their dominance in the private market.

Conclusion

An antimonopoly approach can help mitigate the harms of extreme concentration in the AI sector. By introducing public utility tools such as structural separations, nondiscrimination requirements, and interoperability rules; public options; and pro-competition industrial and procurement policies, policymakers can shape how AI is developed, deployed, and used—and in the process, protect innovation, competition, national security, and the American people.

Endnotes

[1] This policy brief is adapted from Ganesh Sitaraman and Tejas Narechania, *An Antimonopoly Approach to Governing Artificial Intelligence*, VANDERBILT POLICY ACCELERATOR (Oct. 2023).

[2] To the extent there has been, it has largely focused on semiconductor manufacturing, and to a lesser extent, cloud infrastructure provision. But even then, these concerns have not generally been considered in the context of AI specifically. Among the rare works to examine competition aspects of AI are C. Scott Hemphill, *Disruptive Incumbents: Platform Competition in an Age of Machine Learning*, 119 COLUM. L. REV. 1973, 1975–81 (2019); Amba Kak & Sarah Myers West, *AI Now 2023 Landscape: Confronting Technology Power*, AI NOW INSTITUTE (April 11, 2023), https://ainowinstitute.org/2023-landscape.

[3] *See infra* Part I, for a discussion.

[4] Matt Bornstein, Guido Appenzeller & Martin Casado, *Who Owns the Generative AI Platform?*, ANDREESENHOROWITZ (Jan. 19, 2023), https://a16z.com/2023/01/19/who-owns-the-generative-ai-platform/

[5] Kif Leswing, *Meet the $10,000 Nvidia Chip Powering the Race for A.I.*, CNBC, Feb. 23, 2023, https://www.cnbc.com/2023/02/23/nvidias-a100-is-the-10000-chip-powering-the-race-for-ai-.html (noting that Nvidia has 95% market share for machine learning); Zoe Corbyn & Ben Morris, *Nvidia: The Chip Maker that Became and AI Superpower*, BBC NEWS, May 30, 2023, https://www.bbc.com/news/business-65675027 (same); Asa Fitch & Jiyoung Sohn, *The Next Challengers Joining Nvidia in the AI Chip Revolution*, WALL ST. J., July 11, 2023 (12:00am ET), https://www.wsj.com/articles/the-next-challengers-joining-nvidia-in-the-ai-chip-revolution-e0055485 (noting that Nvidia has more than 80% of the market); Wallace Witkowski, *Nvidia 'Should Have At Least 90%' of AI Chip Market with AMD on its Heels,* MARKETWATCH, July 10, 2023 (12:50pm ET), https://www.marketwatch.com/story/nvidia-should-have-at-least-90-of-ai-chip-market-with-amd-on-its-heels-13d00bff (projecting Nvidia will have 90% of the chip market).

[6] Arjun Kharpal, *Two of the World's Most Critical Chip Firms Rally After Nvidia's 26% Share Price Surge*, CNBC, May 25, 2023 (7:14 AM EDT), https://www.cnbc.com/2023/05/25/tsmc-asml-two-critical-chip-firms-rally-after-nvidias-earnings.html

[7] TSMC's market share was estimated at 58.5% in 2022, with runner-up Samsung coming in at 15.8%. Peter Clarke, *TSMC, Globalfoundries Gained as Foundry Market Cooled*, EENEWS, March 13, 2023, https://www.eenewseurope.com/en/tsmc-globalfoundries-gained-as-foundry-market-cooled/.

[8] Saif M. Khan & Alexander Mann, *AI Chips: What They Are and Why They Matter*, CENTER FOR SECURITY AND EMERGING TECHNOLOGY 11 (April 2020).

[9] *Id*.

[10] Kate Tarasov, *ASML is the Only Company Making the $200 Million Machines Needed to Print Every Advanced Microchip. Here's an Inside Look*, CNBC (Mar. 23, 2022, 1:00pm EDT), https://www.cnbc.com/2022/03/23/inside-asml-the-company-advanced-chipmakers-use-for-euv-lithography.html.

[11] Clay Pascal, *Nvidia H100 GPUs: Supply and Demand*, GPU UTILS, July 2023, Updated August 2023, https://gpus.llm-utils.org/nvidia-h100-gpus-supply-and-demand-.

[12] Samuel Nyberg, *Apple Gets Special Treatment Amid Chip Shortage*, MACWORLD (Jun. 22, 2021, 1:01pm PDT), https://www.macworld.com/article/677141/apple-gets-special-treatment-amid-chip-shortage.html

[13] Philip J. Weiser, *The Internet, Innovation, and Intellectual Property Policy*, 103 COLUM. L. REV. 534, 579 (2003) ("In the government's antitrust case against Microsoft, for example, the government submitted evidence of a manager's statement that 'to control the APIs is to control the industry' and established that Microsoft's monopoly rested, in part, on its firm control of its APIs.'").

[14] W. KIP VISCUSI, JOSEPH E. HARRINGTON, JR. & JOHN M. VERNON, ECONOMICS OF REGULATION AND ANTITRUST 82 (4th ed. 2005)

[15] We are indebted to Nick Garcia of Public Knowledge for this example.

[16] Investigation of Competition in Digital Markets: H. Comm. On The Judiciary, 117th Congress, 98-99 (2020) https://www.govinfo.gov/content/pkg/CPRT-117HPRT47832/pdf/CPRT-117HPRT47832.pdf

[17] Investigation of Competition in Digital Markets: H. Comm. On The Judiciary, 117th Congress, 96 (2020) https://www.govinfo.gov/content/pkg/CPRT-117HPRT47832/pdf/CPRT-117HPRT47832.pdf

[18] Andrew Leonard, *Amazon Has Gone From Neutral Platform to Cutthroat Competitor, Say Open Source Developers*, ONEZERO (April 24, 2019) https://onezero.medium.com/open-source-betrayed-industry-leaders-accuse-amazon-of-playing-a-rigged-game-with-aws-67177bc748b7; Jordan Novet, *Amazon Steps Up its Open-Source Game, and Elastic Stock Falls as a Result*, CNBC (March 12, 2019) https://www.cnbc.com/2019/03/12/aws-open-source-move-sends-elastic-stock-down.html; Jordan Novet, *Amazon's Cloud Business is Competing with its Customers*, CNBC (Nov. 30, 2018), https://www.cnbc.com/2018/11/30/aws-is-competing-with-its-customers.html.

[19] Sai Krishna Kamepalli, Raghuram Rajan & Luigi Zingales, *Kill Zone*, NBER Working Paper, https://www.nber.org/papers/w27146.

[20] *See, e.g.*, Karen M. Sutter, John F. Sargent, Jr. & Manpreet Singh, *Semiconductors and the CHIPS Act: The Global Context* 2, CONG. RES. SERV., May 18, 2023, https://crsreports.congress.gov/product/pdf/R/R47558

[21] *See, e.g.*, Nick Merrill & Tejas N. Narechania, *Inside the Internet*, DUKE L.J. (forthcoming 2023).

[22] KAI-FU LEE, AI SUPERPOWERS: CHINA, SILICON VALLEY, AND THE NEW WORLD ORDER 170 (Mariner Books ed., 2021) (2018)

[23] MARTIN GILENS, AFFLUENCE AND INFLUENCE: ECONOMIC INEQUALITY AND POLITICAL POWER IN AMERICA 97–123 (2012); LARRY M. BARTELS, UNEQUAL DEMOCRACY 253-54 (2008); Martin Gilens & Benjamin I. Page, *Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens*, 12 PERSP. ON POL. 564, 573 (2014); LEE DRUTMAN, BUSINESS OF AMERICA IS LOBBYING (2015); ALYSSA KATZ, THE INFLUENCE MACHINE: THE U.S. CHAMBER OF COMMERCE AND THE CORPORATE CAPTURE OF AMERICAN LIFE (2015); KAY LEHMAN SCHLOZMAN, SIDNEY VERBA & HENRY E. BRADY, THE UNHEAVENLY CHORUS 404-11 2012).

[24] Tejas Narechania, *Machine Learning as Natural Monopoly*, 107 IOWA L. REV. 1543 (2022).

[25] MORGAN RICKS, GANESH SITARAMAN, SHELLEY WELTON & LEV MENAND, NETWORKS, PLATFORMS, AND UTILITIES: LAW AND POLICY 8-10 (2022).

[26] ID. at 11-21.

[27] ID. at 28.

[28] For a discussion of this example and others, including a theory of structural separations, see Lina M. Khan, *The Separation of Platforms and Commerce*, 119 COLUM. L. REV. 973 (2019).

[29] *Cf.* William P. Rogerson & Howard Shelanski, *Antitrust Enforcement, Regulation, and Digital Platforms*, 168 U. PA. L. REV. 1911, 1934–36 (2020).

[30] If regulators were to determine that integration of chips and cloud is desirable for effective service provision, then separating chips/cloud from higher levels in the stack would encourage innovation across both layers together—while preserving the innovative potential of competition further up the stack.

[31] RICKS, ET AL., *supra* note 25, at 24-26.

[32] ID. at 24, 26, 29.

[33] ID. at 24.

[34] ID. at 25.

[35] Narechania, *supra* note 24, at 1555.

[36] See TensorFlow Federated: Machine Learning on Decentralized Data, TENSORFLOW, https://www.tensorflow.org/federated (describing one approach to federated learning).

[37] This approach is distinct from the one adopted in Europe via Gaia-X, which predominantly regards federated data storage, for the purposes of complying data localization requirements (e.g., rules that certain personal data be housed in certain locales). By contrast, federated learning can describe an interoperable approach to training, in which multiple applications or users train a single, shared foundation model through an interoperable standard.

[38] RICKS, ET AL., *supra* note 25, at 29-30.

[39] ID.

[40] ID.

[41] GANESH SITARAMAN & ANNE ALSTOTT, THE PUBLIC OPTION 27 (2019).

[42] ID. at 38-40.

[43] E.S. Savas, *An Empirical Study of Competition in Municipal Service Delivery*, 37 PUB. ADMIN. REV. 717 (1977); E.S. Savas, *Intracity Competition between Public and Private Service Delivery*, 41 PUB. ADMIN. REV. 46 (1981).

[44] For a discussion, see our companion paper, Tejas Narechania & Ganesh Sitaraman, *An Antimonopoly Approach to Governing Artificial Intelligence*, Vanderbilt Policy Accelerator (Oct. 3, 2023).

[45] Nikkei Staff Writers, *Japan's METI to Build New Supercomputer to Help Develop AI at Home*, NIKKEI ASIA, July 24, 2023 (15:02 JST), https://asia.nikkei.com/Business/Technology/Japan-s-METI-to-build-new-supercomputer-to-help-develop-AI-at-home.

[46] NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH RESOURCE TASK FORCE, STRENGTHENING AND DEMOCRATIZING THE U.S. ARTIFICIAL INTELLIGENCE INNOVATION ECOSYSTEM: AN IMPLEMENTATION PLAN FOR A NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH RESOURCE 22 (Jan. 2023).

[47] *Id*. at 31.

[48] *Id*. ("This could be made available by leveraging an existing supercomputer or newly procured through a competitive bid process managed by the Operating Entity in consultation with the Steering Committee and relevant advisory boards.").

[49] *CREATE AI Act of 2023*, S. 2714, 118th Cong. § 5603(b) (2023). *See* Press Release, Rep. Anna G. Eshoo, AI Caucus Leaders Introduce Bipartisan Bill to Expand Access to AI Research, July 28, 2023, https://eshoo.house.gov/media/press-releases/ai-caucus-leaders-introduce-bipartisan-bill-expand-access-ai-research.

[50] Ben Gansky, Michael Martin & Ganesh Sitaraman, *Artificial Intelligence is Too Important to Leave to Google and Facebook Alone*, N.Y. TIMES, Nov. 10, 2019.

[51] *CREATE AI Act*, *supra* note 46.

[52] *See, e.g.*, AI.gov, *AI Researchers Portal: Data Resources*, https://www.ai.gov/ai-researchers-portal/data-resources/ (last visited Aug. 21, 2023).

[53] *See* Philippe Aghion, Jing Cai, Mathias Dewatripont, Luosha Du, Ann Harrison & Patrick Legros, *Industrial Policy and Competition*, AM. ECON. J.: MACROECONOMICS 7(4): 1-32 (2015).

[54] CHIPS Act of 2022, Pub. L. No. 117-167, Div. A, § 102, 136 Stat. 1372 (2022).

[55] *Id*.

[56] CHIPS FOR AMERICA, VISION FOR SUCCESS: COMMERCIAL FABRICATION FACILITIES 1, NAT'L INST. STANDARDS & TECH., Feb. 28, 2023, https://www.nist.gov/system/files/documents/2023/02/28/Vision_for_Success-Commercial_Fabrication_Facilities.pdf

[57] Note that the Chips office does appear to want *two* clusters in the United States, but does not commit to those being run by two independent firms. *See* ID.