AI AS A GPT: IMPLICATIONS FOR COMPETITION

AVI GOLDFARB & YUVAL LIDANY*

INTRODUCTION

Artificial intelligence (AI) is emerging as a transformative force in the global economy, with the potential to reshape production processes, create new markets, and enhance productivity across sectors. Its economic significance lies not only in automating existing tasks but also in enabling new capabilities, from diagnostic tools in healthcare to adaptive learning systems in education and advanced analytics in finance.[1] These advances can raise living standards, improve public services, and accelerate scientific discovery. This paper explores reasons why AI markets may remain competitive or tip toward concentration, and what that means for innovation and antitrust.

We argue that for AI to reach its full potential to generate widespread prosperity, a competitive environment is needed. Economic theory identifies

---

* [Authors' footnote]

[1] AJAY AGRAWAL, JOSHUA GANS & AVI GOLDFARB, POWER AND PREDICTION: THE DISRUPTIVE ECONOMICS OF ARTIFICIAL INTELLIGENCE 16, 28–29, 47–49, 74–76, 90–94 (2022).

contrasting forces affecting the relationship between competition and innovation. Arrow argued for a positive relationship between competition and innovation, emphasizing that competition creates incentives to improve quality and reduce costs.[2] In contrast, Schumpeter asserted that market power stimulates innovation by enabling firms to capture the rents from their innovations.[3] The Schumpeterian models emphasize that many innovations have high fixed costs and so market power is needed to generate rents to cover the fixed costs. Scherer noted that both forces are likely at play and proposed an inverted U shaped relation- ship between market concentration and innovation.[4] Empirically, competitive pressure is associated with faster productivity growth and technology adoption.[5] Aghion et al. documents an inverted U relationship, where innovation is maximized when competition is sufficient to spur effort but not so intense as to eliminate returns entirely.[6]

For AI, the forces suggesting competition generates more innovation are particularly salient. AI-related products include the set of upstream inputs, core technologies, and downstream channels required to build and deploy AI tools. There are complementarities between the different layers of this stack. If a small number of technology firms control AI, they might be able to throttle its development or direct the technology in a way that serves their own purposes and not those of the broader society.

Because AI has potential to be a general-purpose technology (GPT),[7] the direction of innovation early on has potential to affect the trajectory of the

---

[2] Kenneth J. Arrow, *Economic Welfare and the Allocation of Resources for Invention, in* THE RATE AND DIRECTION OF INVENTIVE ACTIVITY 609, 619 (Universities-Nat'l Bureau Comm. for Econ. Rsch. ed., 1962).

[3] JOSEPH A. SCHUMPETER, CAPITALISM, SOCIALISM AND DEMOCRACY 127 (Taylor & Francis Grp. 2006) (1942).

[4] Frederic M. Scherer, *Market Structure and the Employment of Scientists and Engineers*, 57 AM. ECON. REV. 524, 530 (1967).

[5] Stephen J. Nickell, *Competition and Corporate Performance*, 104 J. POL. ECON. 724, 724 (1996).

[6] Philippe Aghion et al., *Competition and Innovation: An Inverted-U Relationship*, 120 Q. J. ECON. 701, 701 (2005).

[7] THE ECONOMICS OF ARTIFICIAL INTELLIGENCE: AN AGENDA 4 (Ajay Agrawal, Joshua Gans & Avi Goldfarb eds., 2019); Iain M. Cockburn, Rebecca M. Henderson & Scott Stern, *The Impact of Artificial Intelligence on Innovation: An Exploratory Analysis, in id.* at 115, 118–20; Avi Goldfarb, Bledi Taska & Florenta Teodoridis, *Could machine learning be a general purpose technology? A comparison of emerging technologies using data from online job postings*, 52 RSCH. POL'Y 104653 (2023).

technology well into the future.[8,9] GPTs generate co-invention in an upstream producing sector and in downstream application sectors.[10] If the upstream producing sector is focused on the needs of vertically integrated downstream applications, the broader co-invention cycle between producing and application sectors may never get going.

In this article, we first describe what we mean by artificial intelligence and the literature on whether AI is a GPT. This literature highlights a co-innovation cycle with GPTs that suggests that a vertically integrated dominant player might affect the direction of innovation. We then describe the various layers of the stack in which this innovation might take place.

The bulk of the article then details four distinct aspects of the economics of AI that might affect the nature of competition. These are complementarity between products, sunk costs, diminishing returns to scale in data, and the potential burden of consumer protection regulation.

Complementarity between products creates both opportunities and risks. On the opportunity side, modular architectures, interoperability, well-designed standards, and thoughtfully structured partnerships can lower barriers to entry, coordinate complementary assets (such as compute and foundation models), speed deployment, and deliver direct consumer benefits through better performance and reduced friction. On the risk side, firms can use vertical integration and exclusive partnerships to preference their products or those of their partners, or restrict rivals' access to critical inputs (such as GPUs, proprietary APIs, data, or distribution). This could lock in customers, foreclose entry, and extend market power across product boundaries, for example by leveraging dominance in established markets to gain power in a market for an AI tool, or vice versa.

Companies are spending billions on hardware, datasets, model training, and talent as they develop AI-related products. Much of this expenditure is a sunk cost, in the sense that it cannot be recovered. Sutton

---

[8] Bresnahan & Trajtenberg (1995) define general-purpose technologies by pervasiveness across downstream sectors, inherent potential for technical improvement, and innovational complementarities that raise the productivity of research and development in downstream sectors.

[9] Timothy F. Bresnahan & Manuel Trajtenberg, *General purpose technologies 'Engines of growth'?*, 65 J. ECON. 83 (1995).

[10] Timothy F. Bresnahan & Shane Greenstein, *Technical Progress and Co-Invention in Computing and in the Uses of Computers*, 27 BROOKINGS PAPERS ON ECON. ACTIVITY 1 (1996).

(1991,1998) highlighted that such high sunk costs could create barriers to entry, even in a rapidly expanding market.[11] Specifically, as market leaders compete with each other, the technology improves and barriers to entry grow. When an industry starts, it might take a few million dollars to enter. But after decades of competition, it might require tens of billions. Endogenous sunk costs can lead to oligopoly, and the textbook argument that an expanding market leads to widespread competition may not apply.

Data exhibits diminishing returns to scale: the more data you already have, the less additional information each new data point provides. Combined with the availability of curated or domain-specific data, smaller entrants can compete effectively in targeted ap- plications. At the same time, large, proprietary datasets generate feedback loops that can strengthen early movers, raising barriers to entry. Thus, data advantages only enhance market dominance under specific conditions.

Finally, well-designed consumer protection rules can reduce harms related to privacy, bias, and safety that a competitive environment alone may not resolve. However, compliance burdens can disproportionately affect smaller firms. Regulatory fragmentation across jurisdictions also raises costs that incumbents with global reach can more easily absorb, potentially reinforcing concentration even as rules aim to increase accountability. Overall, the mixed effects of these different aspects of the economics of AI suggest good and bad news for the future of competition in AI-related markets. Modular systems, minimum efficient scale, diminishing returns to scale in data, and thoughtful consumer protection regimes suggest the potential for competitive AI markets. This is consistent with the degree of competition for the next generation of innovative AI tools as of the time of writing in August 2025. Established technology companies building their own internal AI tools include Google, Apple, Amazon, Palantir, and Alibaba. Meta has made some of their models open, presumably hoping to profit from AI's complements. A number of more recently founded organizations have also built large scale and effective AI models, including OpenAI, Anthropic, xAI, Mistral, Cohere, and Deepseek. Many other companies are developing useful and specialized AI tools in a variety of applications, including autonomous driving, banking and financial services, graphic design, and healthcare.

[11] JOHN SUTTON, SUNK COSTS AND MARKET STRUCTURE: PRICE COMPETITION, ADVERTISING, AND THE EVOLUTION OF CONCENTRATION 11–12 (1991).

There are nevertheless substantial risks to the future of competition in AI-related markets. Complementarity between AI and AI-related products could allow companies to tie goods across markets, enabling dominance in some technology markets to create market power in others. As a small number of large players invest in improving their models, it could raise barriers to entry ultimately leading to substantial market power. Data feedback loops may reinforce such advantages as a firm's data collected in one market is used to create advantages in others. The regulatory environment may also create barriers to entry for smaller firms and new innovative ideas.

In the case of endogenous sunk costs and scale in data, which forces dominate depends on the underlying economics. For regulatory burden, it depends on government decisions with respect to legislative tradeoffs. For complementarity between products, it instead might depend on actions within the legal system. Vigorous antitrust attention to illegal ties can reduce the exploitation of market power and encourage widespread benefits to AI innovation.

I. UNDERSTANDING GENERATIVE AI

A. WHAT IS AI

AI is the theory and development of computer systems able to perform tasks normally requiring human intelligence. The main technology underlying the growing AI industry is a form of artificial intelligence called machine learning, which is a branch of computational statistics.[12] AI systems using machine learning can replicate or augment human cognitive capabilities, enabling machines to handle complex tasks with increasing accuracy and efficiency.[13] Economically, they enhance productivity and innovation across industries, from healthcare to finance, by automating some tasks and enhancing human capabilities in others.

Key subfields of AI underpin its diverse applications. Natural language processing (NLP) focuses on enabling machines to interpret, generate, and interact using human language, supporting tasks like translation and dialogue systems. Computer vision allows machines to process and understand visual

---

[12] Ajay Agrawal, Joshua Gans & Avi Goldfarb, *Introduction to The Economics of Artificial Intelligence: An Agenda, in* THE ECONOMICS OF ARTIFICIAL INTELLIGENCE: AN AGENDA, *supra* note 7, at 1, 3.

[13] Erik Brynjolfsson, Daniel Rock & Chad Syverson, *Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics* (Nat'l Bureau of Econ. Rsch., Working Paper No. 24001, 2017).

data, such as images or videos, for applications like object detection. Speech synthesis generates human-like speech from text, enabling virtual assistants and accessibility tools. Autonomous decision-making involves systems selecting optimal actions based on data and predefined goals, as seen in robotics and strategic planning.[14] Recent advancements highlight the development of agentic AI capable of autonomous action and of multimodal systems that fuse multiple sensory domains.[15]

Technically, contemporary AI is dominated by machine learning, in which models learn patterns from data to fill in missing information, enabling software to make predictions, generate content, or choose actions rather than rely on hand-coded, task-specific rules. AI tools often depend on specialized hardware, including graphics processing units (GPUs) and tensor processing units (TPUs); software frameworks (e.g., TensorFlow, PyTorch, transformer libraries); and large, high-quality datasets that materially affect performance in tasks such as translation, sentiment analysis, and object detection.[16]

AI systems are deployed across a wide range of applications, reflecting their versatility. In natural language processing, large language models enable tasks such as automated translation, text generation, and conversational agents.[17] In computer vision, AI powers facial recognition, medical imaging

---

[14] RECOMMENDER SYSTEMS HANDBOOK (Francesco Ricci, Lior Rokach & Bracha Shapira eds., 2nd ed. 2015); ARTIFICIAL INTELLIGENCE: A MODERN APPROACH (Stuart J. Russell & Peter Norvig eds., 3rd ed. 2010).

[15] Yoshua Bengio et al., *International AI Safety Report*, INT'L SCI. REP. ON THE SAFETY OF ADVANCED AI (Jan. 29, 2025), https://doi.org/10.48550/arXiv.2501.17805 [https://arxiv.org/abs/2501.17805].

[16] Martín Abadi et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems* (Mar. 16, 2016), https://doi.org/10.48550/arXiv.1603.04467 [https://arxiv.org/abs/1603.04467]; Adam Paszke et al., *PyTorch: An Imperative Style, High- Performance Deep Learning Library* (Dec. 3, 2019), https://doi.org/10.48550/arXiv.1912.01703 [https://arxiv.org/abs/1912.01703]; Rishi Bommasani et al., *On the Opportunities and Risks of Foundation Models (Aug. 18,* 2021), https://doi.org/10.48550/arXiv.2108.07258 [https://arxiv.org/abs/2108.07258]; COMPETITION AND MKTS. AUTH., AI FOUNDATION MODELS: UPDATE PAPER (2024).

[17] Jacob Devlin et al., *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* (May 24, 2019), https://doi.org/10.48550/arXiv.1810.04805 [https://arxiv.org/abs/1810.04805]; Tom B. Brown et al., *Language Models are Few-Shot Learners* (July 22, 2020), https://doi.org/10.48550/arXiv.2005.14165 [https://arxiv.org/abs/2005.14165].

analysis, and autonomous vehicle navigation.[18] Other domains include robotics, where AI enables precise control in manufacturing, and recommendation systems, which personalize content on platforms like streaming services or e-commerce.[19] These applications demonstrate AI's ability to transform industries by automating tasks, enhancing decision-making, and creating new capabilities, from conversational AI like ChatGPT to diagnostic tools in healthcare.[20]

B. AI AS A GPT

A general-purpose technology is a technology that diffuses widely across industries, im- proves over time, and makes other innovation more productive. Specifically, Bresnahan and Trajtenberg define GPTs by their pervasiveness across downstream sectors, sustained scope for technical improvement, and innovation spillovers that raise the productivity of research and development (R&D) in related activities.[21] GPTs have an outsized impact on productivity growth because of a cycle of complementary innovation, or co-invention, between producing industries and downstream applications.[22] Unlike a one-off productivity shock, GPTs sustain growth: innovations upstream stimulate downstream advances, and vice versa, so productivity continues to rise rather than leveling off.

AI is increasingly recognized as a GPT.[23] It is being used in a large number of industries, there is rapid technological improvement in AI models,

---

[18] Kaiming He et al., *Deep Residual Learning for Image Recognition, in* PROCEEDINGS OF THE 29TH IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION 770 (2016).

[19] RECOMMENDER SYSTEMS HANDBOOK, *supra* note 14.

[20] AGRAWAL, GANS & GOLDFARB, *supra* note 1; Karan Singhal et al., *Large language models encode clinical knowledge*, 620 NATURE 172, 172 (2023).

[21] Bresnahan & Trajtenberg, *supra* note 9, at 83; Goldfarb, Taska & Teodoridis, *supra* note 7.

[22] Bresnahan & Trajtenberg, *supra* note 9; Bresnahan & Greenstein, *supra* note 10.

[23] Erik Brynjolfsson, Daniel Rock & Chad Syverson, *Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics, in* THE ECONOMICS OF ARTIFICIAL INTELLIGENCE: AN AGENDA, *supra* note 7, at 23, 39; Iain M. Cockburn, Rebecca M. Henderson & Scott Stern, *The Impact of Artificial Intelligence on Innovation: An Exploratory Analysis, in* THE ECONOMICS OF ARTIFICIAL INTELLIGENCE: AN AGENDA, *supra* note 7, at 115, 133–39; Manuel Trajtenberg, *Artificial Intelligence as the Next GPT: A Political-Economy Perspective, in* THE ECONOMICS OF ARTIFICIAL INTELLIGENCE: AN AGENDA, *supra* note 7, at 175, 175; Ajay Agrawal, Joshua Gans & Avi Goldfarb, *Introduction to The*

and downstream industries are developing new products, services, and organizational processes that take advantage of AI's capabilities.

The GPT character of AI matters for competition policy. First, the potential economy- wide impact of AI suggests that careful attention to competition in AI-related markets is warranted. Furthermore, potentially common upstream inputs in compute, data, and model application programming interfaces (APIs) can affect a variety of downstream markets that use AI. Openness, portability, and interoperability can increase the return to complementary investment, while foreclosure of key inputs or interfaces can depress innovation across sectors.[24]

These observations motivate our focus on complementarity between products in the AI stack, on endogenous sunk costs affecting the direction of innovation, on data feedback loops, and on consumer protection regulation. In each case, the cycle of AI innovation can amplify anticompetitive harms. If AI is a GPT, early concentration could distort its trajectory across the entire economy.

II. THE AI STACK

Companies that develop and deploy AI operate within a layered stack of interrelated technologies, services, and business models. Each layer plays a distinct role in enabling AI systems and influences how firms compete, collaborate, and consolidate market power. Unlike the product markets described in introductory economics textbooks, AI relies on complementary inputs across layers, such that control at one point in the stack can shape outcomes at others.

Economics of Artificial Intelligence: An Agenda, in THE ECONOMICS OF ARTIFICIAL INTELLIGENCE: AN AGENDA, *supra* note 7, at 1, 6.

[24] Bresnahan & Trajtenberg, *supra* note 9; Bresnahan & Greenstein, *supra* note 10; Aghion, Jones & Jones, *supra* note 23; COMPETITION AND MKTS. AUTH., *supra* note 16, at 14.

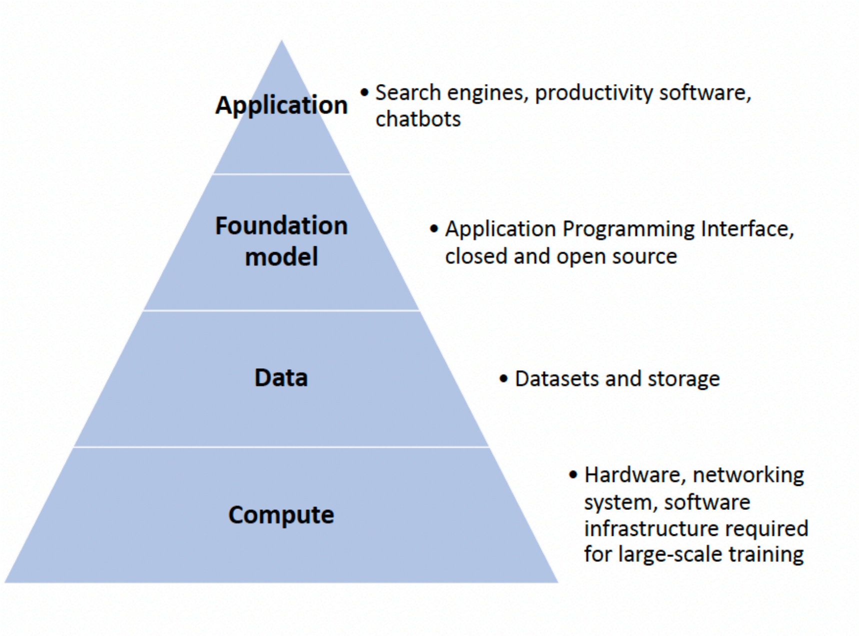

FIGURE 1: THE AI STACK, ADAPTED FROM AZOULAY, KRIEGER, AND NAGARAJ

Azoulay, Krieger, and Nagaraj characterize the AI stack as consisting of four layers.[25,26] At the base is the compute layer, which encompasses GPUs together with the supporting hardware, networking systems, and software

---

[25] Pierre Azoulay, Joshua Krieger & Abhishek Nagaraj, *Old Moats for New Models: Openness, Control, and Competition in Generative Artificial Intelligence, in* 4 ENTREPRENEURSHIP AND INNOVATION POL'Y AND ECON. 7, 12–14 (Benjamin Jones & Josh Lerner eds., 2025).

[26] By contrast, the OECD approaches AI through a five-dimensional classification framework that emphasizes societal and regulatory considerations alongside technical ones. Its framework consists of (i) people and planet, (ii) economic context, (iii) data and inputs, (iv) the AI model, and (v) tasks and outputs. *See* OECD, *OECD Framework for the Classification of AI systems,* OECD DIGIT. ECON. PAPERS, Feb 2022, at 1, 25–54, https://www.oecd.org/en/publications/oecd-framework-for-the-classification-of-ai-systems_cb6d9eca-en.html [https://doi.org/10.1787/cb6d9eca-en]. We adopt Azoulay, Krieger, and Nagaraj as it provides a market-structure-oriented view of AI technologies, directly linking the structure of the stack to potential bottlenecks, competitive dynamics, and antitrust concerns.

infrastructure required for large- scale training. Above this sits the data layer, comprising the vast datasets required for model development and the storage capacity necessary to host them. The third layer is the foundation model, which functions as a general-purpose platform on which down- stream applications are built. Some of these models are closed systems accessible only through APIs, while others provide weight releases that allow developers to inspect and adapt the models for their own purposes. At the top is the application layer, where end users access AI functionality through tools such as search engines, productivity software, or chatbots. This four-layer framework clarifies how competition at individual levels and interdependencies within and across them can create multiple points of competitive leverage.

At present, competitive rivalry is observable at each layer of the stack. In compute, NVIDIA remains by far the largest supplier of high quality GPUs but faces emerging challenges from AMD, Intel, and specialized startups such as Graphcore and Cerebras.[27] In cloud infrastructure, Amazon Web Services, Microsoft Azure, and Google Cloud compete for market share, with Oracle and others providing specialized cloud services.[28] In the model layer, proprietary systems such as OpenAI's GPT-4, Anthropic's Claude, and Google's Gemini compete with weight-released models such as Meta's Llama family and more permissively licensed alternatives such as those from Mistral.[29] At the application layer, established firms such as Microsoft, Google, and Adobe integrate AI into existing product suites, while independent providers such as Jasper and Perplexity AI introduce standalone services. Beyond the technology sector, AI applications diffuse into vertical markets including healthcare, where firms such as PathAI, Deepmind's Alphafold, and Moderna apply AI to diagnostic imaging and drug discovery; finance, where institutions such as JPMorgan, BlackRock, and Mastercard deploy AI for fraud detection, risk management, and credit scoring; education, where companies such as Duolingo, Carnegie Learning, and Khan Academy offer personalized tutoring systems; and logistics, where players such as UPS, FedEx, and Flexport leverage AI for route optimization and

[27] Andrei Hagiu & Julian Wright, *Artificial intelligence and competition policy*, 103A INT'L J. INDUS. ORG. 103134, at 1, 3 (2025).

[28] Gary Biglaiser, Jacques Crémer & Andrea Mantovani, *The economics of the cloud* 1, 2 (Toulouse Sch. of Econ., Working Paper No. 24-1520, 2024); OECD, *Competition in the Provision of Cloud Computing Services,* OECD ROUNDTABLES ON COMPETITION POL'Y PAPERS, NO. 323, at 1, 16 (2025).

[29] Mahyar Habibi, *Open Sourcing GPTs: Economics of Open Sourcing Advanced AI Models* (Jan 20, 2025), https://doi.org/10.48550/arXiv.2501.11581 [https://arxiv.org/abs/2501.11581].

ware- house automation. These patterns suggest that competition is currently dispersed across multiple levels of the stack. Whether this competition endures remains uncertain as training costs escalate and as incumbents strengthen their positions through horizontal tying, vertical integration, and contractual arrangements across layers.

Because AI has the characteristics of a GPT, advances at one layer of the stack both depend upon and reinforce advances at others. This interdependence creates a co- invention cycle: improvements in compute enable larger models, which generate demand for more data, which in turn expand the scope of downstream applications. Competition or market power within a single layer therefore has consequences that extend across the entire stack, shaping not only the set of immediate rivals but also the trajectory of innovation.

This dynamic amplifies the importance of chokepoints. Advances in compute determine the pace of training, while access to data shapes which models can be developed and how they can be adapted to particular uses. Foundation models expand the range of feasible downstream applications, and domain-specific modifications channel their use into specialized markets. Interfaces between layers, such as compute access, model APIs, and data pipelines, play a central role in shaping how competition unfolds. When incumbents control these interfaces, they can limit rivals' ability to scale or interoperate, effectively transforming them into strategic bottlenecks.[30] Market power in AI therefore may stem not only from dominance within a single layer of the stack but also from the strategic use of control points that connect the different layers.

III. ECONOMIC FORCES SHAPING AI MARKET STRUCTURE

In this section, we examine four forces that affect competition within this multi-layered architecture.

A. COMPLEMENTARITY BETWEEN PRODUCTS

AI is not a single product. The AI stack described above highlights the complementarity of different layers, as well as the complementarity of AI with other products. To the extent that AI proves to be a GPT, this complementarity

---

[30] Azoulay, Krieger & Nagaraj, *supra* note 25, at 9; Renee Bohnsack et al., *Profiting from innovation when digital business ecosystems emerge: A control point perspective*, 53 RSCH. POL'Y 10496, at 1 (2024); Hagiu & Wright, *supra* at note 27, at 26–27.

will affect the rate and direction of innovation throughout the economy. The upstream innovation in the AI-producing sectors will affect downstream innovation in AI-application sectors, and downstream innovation will affect processes upstream.

This complementarity creates opportunities for market power in one part of the stack to have an outsized impact on other sectors. Firms can tie products together, or they can choose to build a modular stack. Firms can favor their own products or those of specific partners, instead of allowing open competition. Firms can also create standards that mitigate their ability to tie or treat their own or allied products favorably.

Considering AI as a GPT, complementarities across the stack shape not just market power but the direction of innovation. Advances upstream spur downstream applications, while adoption downstream drives further upstream progress, so foreclosure in one layer can slow innovation across the whole system.

1. *Tying and Modularity*

Complementarity between products means that firms must make decisions about compatibility. Improved compatibility in technology markets often provides consumers with significant utility by enabling enhanced features, streamlined user experiences, and optimized performance.[31] In the context of AI, ensuring compatibility of AI models with cloud services or software platforms might make specific functions such as real-time data processing or tailored user interfaces more effective.

Firms have a choice, however, whether such compatibility is exclusive to their own products or whether to ensure interoperability. Firms can strategically leverage compatibility to tie consumers to their broader product suites. Tying occurs when a firm links two products so customers cannot easily use one without the other. While this can be done with formal bundling or contractual ties, firms can also tie products together through other means. For example, firms can create switching costs and create dependencies that limit their customers' ability to mix and match components from rival providers. This practice of tying distorts market outcomes by reducing interoperability and limiting competitive entry.

---

[31] Michael D Whinston, *Tying, foreclosure, and exclusion*, 80 AM. ECON. REV. 837, 839 (1990).

Rather than tying, firms can instead focus on modularity. In economic theory, modularity refers to the decomposition of complex systems into distinct, interchangeable components that can be independently developed and recombined. This structure can support decentralized innovation by lowering entry costs, encouraging specialization, and facilitating reuse, as long as the cost of building such modular systems is not too high.[32] Lidany provides empirical evidence that tying through compatibility, where technically standalone products are designed to function most effectively when used together, can foreclose rivals even in the absence of formal bundling.[33] Using experimental evidence, the paper demonstrates that by engineering performance advantages across the otherwise stand-alone components of compatible smartphones and laptops, firms replicate the exclusionary effects of traditional tying of add-on products. This encourages adoption of their full suite and discourages substitution to competing products. In this way, tying can extend market power across multiple complementary products.

Building on this experimental evidence, Lidany employs a structural model of com- petition with complementarity between smartphones and laptops. The structural model enables a simulation in which firms must make their laptops function as well with competitors' smartphones as with their own, a type of forced interoperability through modularity. The results suggest that such forced interoperability does increase consumer surplus by enhancing product compatibility and variety. At the same time, forced interoperability can deepen market concentration within each layer when one device is much higher quality than the other. Put differently, interoperability means that consumers are free to choose the best device, and this could increase concentration if most consumers agree on which device is best.

Without forced interoperability with the incumbent, smaller firms can compete only by entering multiple markets and offering compatible, high-quality products across them. This requires substantial investment, raising barriers to entry and limiting the ability of independent firms to challenge incumbents' compatibility advantages.

---

[32] CARLISS Y. BALDWIN & KIM B. CLARK, DESIGN RULES, VOLUME 1: THE POWER OF MODULARITY 63–92 (2000).

[33] Yuval Lidany, Cross-Product Compatibility, Lock-In, and Market Power: The Case of Smartphones and Laptops 1–4 (2025) (unpublished manuscript) (on file at https://yuvallidany.github.io/files/JMP.pdf).

In the AI context, there are a variety of ways that firms can tie or favor their own products, either by leveraging market power in established markets to create market power in AI markets, or vice versa. For example, firms might be able to embed foundation models within broader proprietary systems, creating potential dependencies that shape user behavior and restrict market access. Integration of OpenAI's models into Azure, Google's models into Google Cloud, or Amazon's models into AWS could, under certain conditions, increase market power in both cloud services and AI. Similarly, integration of data from services like Amazon's shopping site, Meta's Instagram, or Google's search pages could increase the appeal of specific AI services if integrated directly into the user interface in a proprietary way.

Modular systems can limit the ability of dominant firms in one market to leverage that market power into complementary markets. For example, modular AI architectures could be built around standardized APIs. Since modularity can promote competition at specific layers, and APIs allow developers to integrate components into broader systems without requiring full system redesign, this could enable smaller firms to participate in various parts of the value chain. There are some early examples of modularity in AI systems, such as the transformer-based APIs provided by OpenAI and Hugging Face.[34]

There are, however, limits to the competitive benefits of modularity. They can encourage competition in some layers of the stack, while reinforcing market power in others. Microsoft's Windows was modular in the sense that other software providers could build on top of it, but the large number of compatible software tools helped reinforce its market power.[35]

This could happen with AI too. To the extent that the application programming interfaces sit on top of proprietary back ends, and pretrained models require specific hardware environments, then a modular software ecosystem can nevertheless reinforce the market power of the providers of hardware in both the chips market and in some adjacent software markets. For example, in addition to the hardware itself, a modular system could favor the software provided by the companies building the hardware chips such as

---

[34] Kevin Boudreau, Lars Bo Jeppesen & Milan Miric, *Artificial Intelligence as a Platform Technology: Strategic Implications of Competing on Top of an AI Platform,* USC MARSHALL SCH. BUS. RSCH. PAPER SERIES 1, 14 (forthcoming July 2025), https://dx.doi.org/10.2139/ssrn.5349634.

[35] United States v. Microsoft Corp., 253 F.3d 34 (D.C. Cir. 2001) (en banc).

NVIDIA, whose GPUs are paired with the proprietary Compute Unified Device Architecture (CUDA) toolkit.[36] Modularity can therefore lower entry barriers in some segments while, in others, reinforcing concentration by consolidating control over essential inputs.

Overall, modularity thus enables decentralized innovation, but control over critical interfaces, whether proprietary APIs, development environments, or hardware tools, can turn it into a gatekeeping mechanism.[37] When dominant firms restrict access to these chokepoints, they can replicate the foreclosure effects of tying while maintaining the appearance of openness. The competitive outcome therefore can hinge on whether modular architectures are governed by open standards or by proprietary control.

2. *Acquisitions and Strategic Alliances*

Complementarity between products creates incentives for acquisitions and strategic alliances. Acquisitions and alliances can promote efficiency by combining complementary assets, accelerating diffusion, and providing exit opportunities that encourage venture investment. They can also facilitate tying and reduce competition.

Once a large firm has acquired a firm with a complementary technology, the concerns above describing how tying might enhance market power apply. After acquisition, it is a single firm that can leverage market power in one product to generate market power in another.[38]

Furthermore, acquisition is not necessary for much of the above discussion of lever- aging market power through tying, and the benefits of modularity, to apply. The ideas extend to strategic alliances between firms. Two complementary products do not need to be provided by the same firm in order for market power in one product to extend to adjacent products.

An important dimension of these alliances is control over critical inputs. In AI, in- puts such as cloud infrastructure, advanced GPUs, and proprietary datasets function as complements to foundation models and applications.

---

[36] COMPETITION AND MKTS. AUTH., *supra* note 16, at 8.

[37] Azoulay, Krieger & Nagaraj, *supra* note 25, at 9.

[38] A separate concern with acquisitions occurs when incumbents acquire startups primarily to neutralize emerging threats. These so-called "killer acquisitions" can reduce the likelihood of follow-on innovation and may entrench dominance by removing independent challengers. *See* Colleen Cunningham, Florian Ederer & Song Ma, *Killer Acquisitions*, 129 J. POL. ECON. 649, 649 (2021).

Strategic partnerships can secure privileged access to these inputs, giving allied firms an advantage in scaling products and reaching customers.[39] Conversely, exclusion from these inputs can limit independent rivals' ability to compete, even if their technologies are otherwise superior. Thus, the role of inputs magnifies the competitive significance of alliances, since restricting or preferentially allocating inputs can replicate many of the same tying and foreclosure concerns associated with acquisitions.

AI commercialization has seen several strategic alliances, such as Microsoft-OpenAI and Amazon-Anthropic. If alliances like this lead to explicit ties between complementary products, then the insights from Whinston, and the related empirical evidence from Lidany, could apply.

Hagiu and Wright formalize this intuition by showing how high concentration in these upstream layers, combined with vertical alliances, could create barriers to entry across the stack.[40] Exclusive partnerships that restrict access to essential inputs deepen dependencies and limit choice. Schrepel and Pentland discuss pro- and anti-competitive reasons for AI foundation model partnerships. Pro-competitive reasons involve enhancing model quality through learning curves, ecosystem effects, or unique data access while anti-competitive reasons foreclose rivals without comparable quality gains, particularly through control over cloud and hardware resources.[41] Korinek and Vipra argue that if there are economies of scale in AI foundation models, then concentrated hardware markets and exclusive alliances could drive vertical integration and diminish competitive opportunities for independent firms.[42]

To our knowledge, there is no rigorous empirical research that has measured whether such alliances have led to market power. It might also be too early for any such research to come to a definitive conclusion. Therefore, the connection between alliances involving AI firms and market power remains an open question.

---

[39] Hagiu & Wright, *supra* at note 27, at 13–14; Anton Korinek & Jai Vipra, *Concentrating intelligence: scaling and market structure in artificial intelligence*, 40 ECON. POL'Y 225, 236–37 (2025).

[40] Hagiu & Wright, *supra* at note 27, at 27–28.

[41] Thibault Schrepel & Alex 'Sandy' Pentland, *Competition between AI foundation models: dynamics and policy recommendations*, 34 INDUS. AND COR. CHANGE 1085, 1096–97 (2025).

[42] Korinek & Vipra, *supra* note 39, at 225.

Table 1: Key cited literature on complementarity between products

Reference	Key Insight	AI Example/ Application	Antitrust Implication
Whinston (1989)	Exclusionary practices can foreclose rivals in interconnected markets (Theory).	Extended to AI via input control in partnerships.	Potential violation of Sherman Act Section 2.
Lidany (2025)	Exclusive compatibility can lead to lock-in and foreclosure.	AI software integration that discourages substitution.	Potential benefits of interoperability mandates.
Korinek & Vipra (2025)	Economies of scale and alliances can drive tipping and vertical integration.	NVIDIA GPU allocations to preferred hyperscalers.	Potential reduction in opportunities for independents.
Hagiu & Wright (2025)	Concentration in hardware/cloud can amplify barriers across the AI stack.	Vertical alliances like AWS–NVIDIA.	Potential for risks to competition.
CMA (2024)	Selective resource allocation can hurt smaller firms.	H100/Blackwell GPU shortages for non-partners.	Potential for exclusionary conduct.

3. *Standards*

Open standards for AI development can facilitate interoperability, supporting the construction of applications that operate across hardware and software systems.[43] These features can reduce vendor lock-in and allow firms to specialize at different layers of the stack, including algorithm design, fine-tuning, or deployment.

More generally, standards can encourage modularity. This makes it more difficult for firms with market power in one product to leverage that market power to create dominance or market power in complementary products.

However, as with modularity, firms with market power over a product that becomes a critical standard can benefit from their market power for

[43] Bommasani et al., *supra* note 16; COMPETITION AND MKTS. AUTH., *supra* note 16; Boudreau, Jeppesen & Miric, *supra* note 34, at 11–12.

that product. It is therefore of strategic value to firms for standards to include their internally developed protocols. As a result, firms try to influence standard setting bodies in their favor. Evidence from consensus standard setting shows that commercialization and distributional conflict slow decision making. Simcoe documents how firms joined internet standard setting bodies in ways that suggest rent seeking. The associated conflict lengthened the time to standard and dampened outcomes likely to upset the status quo.[44] This empirical evidence builds on theory by Farrell and Simcoe, who model consensus selection as a war of attrition in which delays can screen for quality when sponsors privately observe proposal value.[45] This work suggests ways to make the process more efficient, including consensus, random selection, and neutral intervention, in order to avoid stalemates and efficiently reflect vested interests and differences in proposal quality. Firms that contribute to widely used interoperability projects or industry forums can influence choices about interfaces, testing protocols, and audit practices that ultimately fit their cloud or software environments.[46] Participation by large providers in standard setting, or in privacy and safety initiatives, can promote important forms of interoperability while simultaneously reinforcing reliance on their infrastructure.

For AI safety and interoperability forums, participation rules, voting thresholds, dis- closure obligations, and licensing terms are therefore not neutral. Time limits, neutral chairs, and open licensing can curb strategic delay. In this way, standards can lower switching costs and entry barriers without inefficiently enabling the market power of firms in any particular layer of the stack to generate market power in complementary layers.

B. SUNK COSTS

Training GPT-4 is estimated in the tens of millions; xAI's Colossus supercomputer cost $7 billion.[47] Economists distinguish between fixed costs

---

[44] Timothy Simcoe, *Standard setting committees: Consensus governance for shared technology plat- forms*, 102 AM. ECON. REV. 305, 305 (2012).

[45] Joseph Farrell & Timothy Simcoe, *Choosing the rules for consensus standardization*, 43 RAND J. ECON. 235, 235 (2012).

[46] JACQUES CRÉMER, YVES-ALEXANDRE DE MONTJOYE & HEIKE SCHWEITZER, COMPETITION POLICY FOR THE DIGITAL ERA (Directorate-General for Competition, European Commission, 2019); COMPETITION AND MKTS. AUTH., *supra* note 16, AT 15–16.

[47] Ben Cottier et al., *The Rising Costs of Training Frontier AI Models* (May 31, 2024), https://doi.org/10.48550/arXiv.2405.21015 [https://arxiv.org/abs/2405.21015]; Konstantin F. Pilz et al., *Trends in AI*

(large one-time investments that do not vary with output) and sunk costs (irrecoverable expenditures). What matters for market structure is that in AI, many of the most important investments fall into the latter category, creating potential barriers to entry. Such sunk costs include the design and validation of specialized accelerator chips, the construction or leasing of data centers and custom chips, prolonged training runs on thousands of GPUs, the acquisition and curation of large proprietary datasets, and the development or licensing of APIs. While certain small-scale models or limited-purpose infrastructure can be built with modest resources, meaningful participation at the technological frontier typically requires substantial and irreversible commitments.

Standard economic models highlight the role of minimum efficient scale, defined as the smallest output level at which a firm achieves the lowest long-run average cost, in determining market structure. If the minimum efficient scale is large relative to market demand, only one or a few firms can profitably operate, raising risks of monopolization. By contrast, if the minimum efficient scale is small relative to demand, concentration should fall as markets expand, since more firms can cover the fixed cost of entry.[48] In this framework, high fixed or sunk costs alone do not imply persistent concentration when markets are sufficiently large. This reasoning underlies the classic view that markets for general-purpose technologies should support multiple firms: as demand grows across diverse applications, the fixed costs of building the technology can be spread broadly, encouraging entry. This logic would suggest that as AI diffuses into search, productivity software, healthcare, and other markets, concentration should diminish over time. Yet the experience of other technology industries shows that this logic does not always hold, because sunk costs can escalate with rivalry.[49]

Sutton emphasizes that the prediction of declining concentration as markets expand may not hold when sunk costs are endogenous rather than exogenous and unchanging. Endogenous sunk costs are expenditures that escalate with rivalry, such as advertising, quality improvements, or continuous research and development, and are strategically chosen by firms to strengthen their competitive position. In such industries, firms repeatedly reinvest to improve their products, so the equilibrium level of sunk costs rises with market size. As a result, concentration often does not fall in large

---

supercomputers (Apr. 23, 2025), https://doi.org/10.48550/arXiv.2504.16026 [https://arxiv.org/abs/2504.16026].

[48] JOHN SUTTON, *supra* note 11.

[49] *See id.*

markets but instead remains stable or even increases, since only firms able to sustain these escalating expenditures can remain competitive.

AI markets exhibit many features of an endogenous sunk cost industry. Leading firms continually retrain models, expand parameter counts, acquire proprietary datasets, and upgrade infrastructure to match or exceed rivals' performance. Sevilla et al. document that training costs for frontier models doubled about every six months from 2012 to 2022, reflecting this rapid escalation.[50] Consistent with Sutton's framework, incumbents such as OpenAI, Google, and Meta have committed hundreds of millions of dollars to foundation model development,[51] investments that are unrecoverable and must be repeated to remain at the frontier. These expenditures can disadvantage firms without access to capital, infrastructure, or elite research talent.[52] Thus, even if AI proves to be a GPT with a very large market, endogenous sunk costs may limit the number of competitors at various layers of the stack.

Some of these expenditures involve talent for AI development. Industry salaries far exceed academic or startup norms, and leading firms attract elite researchers by offering privileged access to proprietary datasets and compute resources.[53] Sutton emphasized R&D expenses as sunk costs. The salaries given to top AI researchers by well-resourced firms may accelerate productivity and commercialization, but they are also part of the sunk cost of R&D.[54]

---

[50] Jaime Sevilla et al., *Compute Trends Across Three Eras of Machine Learning*, *in* 2022 INT'L JOINT CONF ON NEURAL NETWORKS (IJCNN), INST. ELEC. AND ELEC. ENG'RS, July 2022, at 1, 4.

[51] Schrepel & Pentland, *supra* note 41, at 1089.

[52] Ajay Agrawal, Joshua Gans & Avi Goldfarb, *Artificial Intelligence and International Trade, in* THE ECONOMICS OF ARTIFICIAL INTELLIGENCE: AN AGENDA, *supra* note 7, at 463, 468–70, 475; Iain M. Cockburn, Rebecca M. Henderson & Scott Stern, *The Impact of Artificial Intelligence on Innovation: An Exploratory Analysis, in* THE ECONOMICS OF ARTIFICIAL INTELLIGENCE: AN AGENDA, *supra* note 7, at 115, 142.

[53] Berber Jin, Keach Hagey & Ben Cohen, *Mark Zuckerberg Leads Meta's Effort to Poach AI Talent from Rivals*, WALL STREET J. (July 18, 2025, 20:00 ET), https://www.wsj.com/tech/ai/meta-ai-recruiting-mark-zuckerberg-sam-altman-140d5861; Robert Cyran, *AI "Lottery Ticket" Scramble Sparks an R&D Spiral*, REUTERS (July 25, 2025, 08:00 EDT), https://www.reuters.com/default/ai-lottery-ticket-scramble-sparks-an-rd-spiral-2025-07-24/.

[54] A separate risk of top talent going to industry relates to what Jurowetzki et al. call the "privatization of AI research". If research is less open, it may be harder for teams

A further implication of endogenous sunk costs is that minimum efficient scale in AI may not be static but cascading. Posner and Weyl argue that in some industries, scale is not only a source of efficiency but also a prerequisite for qualitative improvements, as higher levels of scale enable new products or functionalities by creating successive thresh- olds rather than a single fixed point.[55] Discussions of 'AI scaling laws' by Kaplan et al., Wei et al., and others illustrate this dynamic: emergent capabilities in large language models, such as coherent text generation, multilingual reasoning, and coding assistance, appear only after parameter counts and training data cross critical thresholds.[56] These dynamics show why Sutton's description of the classic prediction of declining concentration with market growth may not apply in AI: because each wave of scaling may create new thresholds that demand escalating investment.

At the same time, countervailing forces exist that can lower effective sunk costs, provide entrants with alternative paths, and enhance competition. Open-source communities such as Hugging Face and platforms such as GitHub host thousands of models, datasets, and libraries that developers can freely use, adapt, and contribute to, enabling broad participation across the AI landscape.[57] Stability AI's release of Stable Diffusion weights, based on the latent diffusion architecture of Rombach et al., has since been ex- tended by a wide community of researchers and smaller firms.[58] Commercial providers such as MosaicML offer optimized training platforms and curated datasets

---

outside the top AI labs to build on proprietary innovation. This in turn may slow the co-invention feedback loop central to GPT's impact on productivity. *See* Roman Jurowetzki et al., *The Privatisation of AI Research—Causes and Potential Consequences* (Feb. 15, 2021), https://doi.org/10.48550/arXiv.2102.01648 [https://arxiv.org/abs/2102.01648].

55 ERIC POSNER & ERIC WEYL, RADICAL MARKETS: UPROOTING CAPITALISM AND DEMOCRACY FOR A JUST SOCIETY 226–30 (2018).

56 Jared Kaplan et al., *Scaling Laws for Neural Language Models* (Jan. 23, 2020), https://doi.org/10.48550/arXiv.2001.08361 [https://arxiv.org/abs/2001.08361]; Jason Wei et al., *Emergent Abilities of Large Language Models* (Oct. 26, 2022), https://doi.org/10.48550/arXiv.2206.07682 [https://arxiv.org/abs/2206.07682].

57 Thomas Wolf et al., *Transformers: State-of-the-art natural language processing, in* PROCEEDINGS OF THE 2020 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING: SYSTEM DEMONSTRATIONS 38, 38 (Qun Liu & David Schlangen eds., 2020), https://doi.org/10.18653/v1/2020.emnlp-demos.6.

58 Robin Rombach et al., *High-Resolution Image Synthesis with Latent Diffusion Models, in* 2022 IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR) 10674, 10674 (2022), https://doi.org/10.1109/CVPR52688.2022.01042.

that reduce per-run expenses, while cloud service providers such as AWS, Azure, and Google Cloud allow firms to rent high-performance chips instead of investing in costly dedicated infrastructure. Advances in parameter-efficient architectures, such as sparse models and mixture-of-experts designs, provide another path to strong performance without extreme scale. Similarly, transfer learning and low-rank adaptation techniques enable entrants to build competitive systems by fine-tuning existing large models rather than training from scratch, while access to domain-specific datasets can allow smaller models to outperform larger systems on targeted tasks.

In summary, high sunk costs alone do not suggest natural monopoly or market power in large and rapidly expanding markets. However, if competition between a small number of firms leads to increased investment, Sutton demonstrated that such endogenous sunk costs can lead to concentrated markets and market power. These endogenous sunk costs can be reinforced to the extent that new capabilities emerge as models scale. Open systems and innovations that reduce the need for scale can push against these forces. While endogenous sunk costs may lead to substantial market power at various layers of the AI stack, empirically it remains an open question.

C. DIMINISHING RETURNS TO SCALE IN DATA

Data is a central input in the production of AI. Prediction accuracy reflects the joint contribution of data, model design, and computational resources.[59] Holding the model class and resources fixed, the marginal value of additional data depends on how much it reduces uncertainty relative to existing information and on the decisions it informs.[60]

---

[59] Susan Athey, *The Impact of Machine Learning on Economics, in* THE ECONOMICS OF ARTIFICIAL INTELLIGENCE: AN AGENDA, supra note 7, at 507, 507–09; Susan Athey & Guido W. Imbens, *Machine Learning Methods That Economists Should Know About*, 11 ANN. REV. ECON. 685, 609–91 (2019); AJAY AGRAWAL, JOSHUA GANS, AND AVI GOLDFARB, PREDICTION MACHINES: THE SIMPLE ECONOMICS OF ARTIFICIAL INTELLIGENCE 32 (2018); Hal Varian, *Artificial Intelligence, Economics, and Industrial Organization, in* THE ECONOMICS OF ARTIFICIAL INTELLIGENCE: AN AGENDA, *supra* note 7, at 399, 400; Ajay Agrawal, Joshua Gans & Avi Goldfarb, *Introduction to The Economics of Artificial Intelligence: An Agenda, in* THE ECONOMICS OF ARTIFICIAL INTELLIGENCE: AN AGENDA, *supra* note 7, at 1, 12.

[60] LAURA L. VELDKAMP, INFORMATION CHOICE IN MACROECONOMICS AND FINANCE 11–12, 86–91 (2011).

As information accumulates, on average each additional piece of data becomes less useful.[61] The first million data points are far more valuable than the billionth. Additional data improves accuracy, but at a slowing rate. This is the idea behind the Central Limit Theorem in statistics. This theory is empirically relevant in the context of digital data and AI. Marginal returns to additional data diminish as datasets grow large. For instance, Bajari et al. show that forecast error falls with data size at a square-root rate.[62] Similarly, Kaplan et al. finds that language-model performance scales with data up to a point, after which accuracy gains plateau and further improvements require better model architectures or higher-quality data.[63] Put differently in a direct sense, as long as there is sufficient data for the model to be useful, there are diminishing returns to scale in data.

A second force suggesting that data as an input enhances the competitive environment is that the economics of data are shaped by its nonrival nature. As Jones and Tonetti emphasize, data can be used by any number of agents simultaneously without diminishing its value.[64] This property creates the potential for large social gains when data is widely shared and reused across applications.[65] Broad access can accelerate innovation, reduce the need for repeated data collection, and improve productivity.

While these two forces suggest that data may not lead to market power, there are situations in which data as a critical input into AI systems can enhance market concentration. For example, data can generate increasing returns through self-reinforcing feedback loops. Data can enable performance gains, which attract additional usage, which in turn can generate additional data. This can create a durable advantage that is difficult for later entrants to overcome.[66] Models of the data economy formalize how information acquisition and data accumulation create dynamic complementarities that

---

[61] *See id*; Maryam Farboodi & Laura Veldkamp, *A Model of the Data Economy* 1 (Nat'l Bureau of Econ. Rsch., Working Paper No. 28427, 2021).

[62] Patrick Bajari et al., *The Impact of Big Data on Firm Performance: An Empirical Investigation*, 109 AEA PAPERS & PROCS. 33, 34 (2019).

[63] Kaplan et al., *supra* note 56.

[64] Charles I. Jones & Christopher Tonetti, *Nonrivalry and the Economics of Data*, 110 AM. ECON. REV. 2819, 2819 (2020).

[65] Hal Varian, *Artificial Intelligence, Economics, and Industrial Organization, in* THE ECONOMICS OF ARTIFICIAL INTELLIGENCE: AN AGENDA, *supra* note 7, at 399, 405; Avi Goldfarb & Catherine E. Tucker, *Digital Economies* 57 J. ECON. LITERATURE 3, 12–16 (2019).

[66] Ajay Agrawal, Joshua Gans & Avi Goldfarb, *How to Win with Machine Learning*, Harv. Bus. Rev., Sept.–Oct. 2020, at 126, 131.

raise effective entry barriers as experience grows.[67] For instance, Google's search may have benefited from such feedback loops, in which better quality leads to more users, better data, and further enhanced quality.[68] Feedback loops may also apply to digital advertising.[69] Ginart et al. provides a formal account of how higher-performing predictors attract more usage, leading to greater data accumulation, faster improvement, and market segmentation as firms specialize in areas of existing strength.[70] AI models may prove to be similar, for example if Google's integration of user interactions from Search, YouTube, or Maps allow it to refine models, personalize responses, and reinforce engagement, and if this in turn leads to sufficiently high quality to attract more users.

To the extent that data is a strategic asset that confers competitive advantage, its non-rival nature may not lead to data's widespread diffusion. Firms can choose not to make data public. While this creates incentives to invest in data, it also constrains the degree to which non-rivalry affects competition.

Overall, data is a critical input into AI products. This does not directly lead to concentrated markets and market power, as there are decreasing returns to scale in data in a technical sense and data is non-rival. At the same time, feedback loops imply that firms with an initial quality advantage can create sustained advantage. Combining the feedback loop advantage with endogenous sunk costs further suggests that barriers to entry into AI-related markets may grow over time, even as the markets themselves get bigger.

D. Regulatory Burden of Consumer Protection

Consumer protection rules are essential for addressing harms such as privacy, bias, and safety, but they can also reshape market structure by raising barriers to entry.

---

[67] Farboodi & Veldkamp, *supra* note 61; Laura Veldkamp & Cindy Chung, *Data and the Aggregate Economy*, 62 J. ECON. LITERATURE 458, 458 (2024).

[68] Agrawal, Gans & Goldfarb, *supra* note 66, at 132.

[69] FIONA M. SCOTT MORTON & DAVID C. DINIELLI, ROADMAP FOR A DIGITAL ADVERTISING MONOPOLIZATION CASE AGAINST GOOGLE 1–3 (2020); Joshua S. Gans, *Market Power in Artificial Intelligence* 1, 23 (Nat'l Bureau of Econ. Rsch., Working Paper 32270, 2024).

[70] Antonio A. Ginart et al., *Competing AI: How does competition feedback affect machine learning?, in* PROCEEDINGS OF THE 24TH INT'L CONF. ON A.I. AND STAT. 1693, 1693 (Arindam Banerjee & Kenji Fukumizu eds., 2021).

Consumer protections for privacy have already affected competition. For example, Campbell, Goldfarb, and Tucker argue that several features of privacy regulation can advantage large multiproduct firms.[71] Evidence from the EU's GDPR is consistent with this prediction. Janßen et al. find that the European Union GDPR reduced mobile app offerings from smaller developers without substantially affecting the larger ones, likely because fixed compliance costs weigh more heavily on smaller players.[72] The GDPR also likely strengthened the largest advertising platforms.[73] This means rules designed to protect consumers may unintentionally protect incumbents if compliance costs fall hardest on small entrants.

AI regulation may also lead to a tradeoff between consumer protection and competition. For example, the European Union AI Act imposes several requirements on high risk systems, including documentation, auditability, and human oversight, which favor firms with extensive legal and technical resources.[74] These costs can amplify sunk cost barriers, as startups struggle to comply, further entrenching market concentration.[75] In this way, the asymmetric impact of regulation can raise entry barriers for smaller firms.[76]

Regulatory fragmentation adds another layer. If different states and different countries have non-nested regulatory frameworks, then the costs of compliance grow. This could further increase the regulatory burden and make it more difficult for smaller firms to compete. In contrast, incumbents with global reach can absorb such costs more easily. Where regulation is widely recognized as necessary, national or global coordination can reduce the disproportionate impact of regulatory compliance on smaller and newer firms. In summary, there is often a tradeoff between aspects of consumer protection and competition. This tradeoff has been documented in the context of privacy regulation, and there are reasons to anticipate that it may also emerge in AI regulation.

---

[71] James Campbell, Avi Goldfarb & Catherine Tucker, *Privacy Regulation and Market Structure*, 24 J. ECON. & MGMT. STRATEGY 47, 47 (2015).

[72] Rebecca Janßen et al., *GDPR and the Lost Generation of Innovative Apps* 1 (Nat'l Bureau of Econ. Rsch., Working Paper 30028, 2022).

[73] Garrett A. Johnson, Scott K. Shriver & Samuel G. Goldberg, *Privacy and Market Concentration: Intended and Unintended Consequences of the GDPR*, 69 MGMT. SCI. 5695, 5696 (2023).

[74] COMPETITION AND MKTS. AUTH., *supra* note 16, at 22–23.

[75] Jones & Tonetti, *supra* note 64; Cockburn, Henderson & Stern, *supra* note 52.

[76] Crémer, Montjoye & Schweitzer, *supra* note 46.

CONCLUSION

AI has potential to be the next GPT, with an outsized impact on productivity and economic growth. GPTs have such a large impact on the economy through a co-invention cycle between complementary products. This process of co-invention means that market power in one part of the stack can affect the rate and direction of innovation throughout. We identified four economic forces that shape competition and innovation in AI. First, we discussed how complementarity between different AI products creates incentives to tie products together, thereby reinforcing market power. Modularity and standards can push against these incentives. Second, we discussed how endogenous sunk costs increase barriers to entry as innovation accumulates. In AI, the increasing costs of building frontier systems suggest endogenous sunk costs are already at play. Third, we discussed the role of data. While there are diminishing returns to scale in data, feedback loops between more users, better data, better quality, and then more users can enable sustained market power. Fourth, AI comes with risks to consumers that have led to consumer protection regulations. When such regulations increase costs of entry, they hurt smaller and newer firms relative to large and established ones.

AI's impact as a GPT will depend not only on technology but also on whether markets remain open to competition. However, it is still an open question whether various AI-related markets will be sufficiently competitive to constrain market power and generate widespread benefits to innovation. This partially depends on underlying economic forces related to endogenous sunk costs and diminishing returns to scale in data. It also partially depends on a policy tradeoff between certain aspects of consumer protection and competition.

For antitrust practitioners, it depends on enforcement with respect to tying and policies that mandate modularity. While sunk costs and data scale reflect the underlying economic structure, tying and modularity depend on legal enforcement. Therefore, antitrust choices are central to AI's future.