IS DATA REALLY A BARRIER TO ENTRY?
RETHINKING COMPETITION REGULATION IN GENERATIVE AI

Alden Abbott, *Mercatus Center*
Satya Marar, *Mercatus Center*

ABSTRACT

The rise of generative artificial intelligence (AI) technology has raised concerns that a few large companies or digital platforms will corner the market. By controlling vast troves of data necessary to train foundation models, these firms could raise entry barriers, stifle competition and innovation, and ultimately harm consumers. Contrary to this fear, thriving data markets allow developers to access training data on an open-source or commercial basis, making it so that data access is not currently a barrier to entry in the generative AI market. In fact, tech giants with access to high volumes of user data face stiff competition from independent firms and startups that can outcompete larger rivals through superior algorithms or user experiences. Technological advancements, such as synthetic data and engineering workarounds that lower data demands and training costs, also suggest that concerns about data scarcity limiting AI innovation are overblown. Additionally, the industry's shift towards more specialized foundation models and tailored AI applications also demonstrates that access to the "right" kind of data rather than vast troves of user data is paramount, further eroding the purported advantage of large tech giants. Conversely, preemptive regulations and remedies proposed to combat future data scarcity or monopolization would likely harm AI development and innovation. Current US antitrust law instead provides a flexible, pragmatic framework for policing exclusionary conduct by firms controlling key data inputs should these concerns materialize in the future.

AUTHOR CONTACT INFORMATION

Alden Abbott, Mercatus Center at George Mason University, aabbott@mercatus.gmu.edu

Satya Marar, Mercatus Center at George Mason University, smarar@mercatus.gmu.edu

Is Data Really a Barrier to Entry?

Rethinking Competition Regulation in Generative AI

The rise of generative artificial intelligence (AI) has prompted global competition authorities to raise concerns that the sector is highly susceptible to monopolization. They argue that large digital platforms benefit from economies of scale and network effects advantages, giving these platforms a significant competitive advantage that could lead to anticompetitive outcomes through their control over data. This understanding has prompted authorities to suggest preemptive antitrust interventions. Our research indicates not only that these concerns are unjustified and overstated but that preemptive antitrust interventions could harm innovation and competition.

Generative AI foundation models (FMs) use machine learning software algorithms to ascertain and predict statistical relationships between data point inputs in order to produce outputs, such as answers to questions, video, audio, images, and more.[1] These models typically rely on and are refined by[2] training on large datasets consisting of billions or even trillions of parameters.[3] [4] Proponents of preemptive antitrust intervention in generative AI, including the US Federal Trade Commission (FTC),[5] theorize that a few firms with access to large troves of proprietary and user data or the resources to acquire that data[6] may develop an unassailable advantage in generative AI technology. They may then use this advantage to curtail competition and innovation by restricting their rivals' access to data[7] and by charging users higher prices without being disciplined by competitive pressure.[8]

---

[1] See: UK Competition and Markets Authority, "AI Foundation Models: Initial Report," September 18, 2023.

[2] Pretraining typically involves feeding a foundation model diverse and large datasets to teach it about general patterns and relationships between data points and types. Using fine-tuned data further adapt a pretrained foundation model to more rapidly or accurately perform specific downstream tasks, typically those of specific applications built upon the foundation model's infrastructure. Applications and models may also be refined through "reinforcement-based learning," whereby human users or programmers rank the outputs that an AI application or model produces in response to specific queries, and these rankings allow the model or application to predict better answers to similar queries in the future. See Matt Mittelsteadt, "Artificial Intelligence: An Introduction for Policymakers" (Mercatus Policy Research, Mercatus Center at George Mason University, Arlington, VA, February 16, 2023); Daniel M. Ziegler et al., "Fine-Tuning Language Models from Human Preferences," *arXiv* E-PRINTS (Jan. 8, 2020), https://arxiv.org/abs/1909.08593; Zhiqing Sun et al., "Aligning Large Multimodal Models with Factually Augmented RLHF," *arXiv* EPRINTS, Sept. 25, 2023, https://arxiv.org/abs/2309.14525.

[3] Ibid.

[4] For instance, Meta's Llama 3 foundation model was trained on 15 trillion tokens of pretraining data. See "Introducing Meta Llama 3: The Most Capable Openly Available LLM to Date," META, Apr. 18, 2024, https://ai.meta.com/blog/meta-llama-3/.

[5] Comment of US Federal Trade Commission to the US Copyright Office, Artificial Intelligence and Copyright, Docket No. 2023-6 (Oct. 30, 2023), 4, *available at* https://www.ftc.gov/legal-library/browse/advocacy-filings/comment-federal-trade-commission-artificial-intelligence-copyright.

[6] Collecting and curating the datasets used to pretrain foundation models is typically expensive. Refining or "fine tuning" models for specific applications using additional user data or specialized datasets is typically less expensive. See Rishi Bommasani et al., "On the Opportunities and Risks of Foundation Models," *arXiv* preprint arXiv:2108.07258 (2021). Scale economies and capital requirements can have anticompetitive implications and lead to reduced consumer welfare if such costs impose restrictions on market entry. See: Preston McAfee et al., "What Is a Barrier to Entry?," *American Economic Review* 94 (2004).

[7] "[A]cquiring data that helps facilitate matching, sorting, or prediction services may enable the platform to weaken rival platforms by denying them that data." See: US Federal Trade Commission and Department of Justice, Merger Guidelines (2023) at 25, https://www.ftc.gov/system/files/ftc_gov/pdf/2023_merger_guidelines_final_12.18.2023.pdf.

[8] Valéria Faure-Muntian, "Competitive Dysfunction: Why Competition Law Is Failing in a Digital World," The Forum Network, February 24, 2021, https://www.oecd-forum.org/posts/competitive-dysfunction-why-competition-law-is-failing-in-a-digital-

Concerns about large firms and digital platforms' exclusive access to data are also driven by fears about future data scarcity. Increasingly powerful and versatile foundation models designed to perform complex applications are expected to require higher volumes of data and access to increasingly specific kinds of data to further improve.[9] In addition to larger and more diverse datasets, a larger and more diverse set of human feedback (such as user rankings of application or model outputs) may also be required to fine-tune larger, increasingly complex models.[10] Although many leading FMs were trained on publicly available data scraped from the internet, a high proportion of publicly available data has already been used in training,[11] and an increasing proportion of public data now consists of AI-generated outputs that can degrade model quality when used in training.[12]

Antitrust regulators like the FTC theorize that data scarcity will increase the advantage that large tech firms and their digital platforms have over rival AI model providers and developers due to those firms' access to vast volumes of training data from users of their other services.[13] Regulators argue that this exclusive access to data would increase tech giants' ability to harm consumers and would stifle innovation by restricting competition, thereby further entrenching their monopoly.[14]

Contrary to these fears, developers of FMs and the applications that use them have access to a range of data sources, including a competitive market of open-source and commercially available proprietary datasets that are open to all. Similarly, data owners have an incentive to sell their data to big and small firms alike. The history of developing digital platforms and generative AI also demonstrates that the advantages that a firm enjoys from having a large existing database of user data is generally superseded when rivals have superior software and algorithms, thus indicating that access to data is not an 'unassailable advantage' over the competition. The right kind of data for a model or application's intended function is generally more important than raw data volume. Future models and applications will thus likely have more specific uses and so will be trained more efficiently with fewer data. Advancements in the ability to create and correct synthetic data also promise to help to meet data volume demands.

Despite the failure of their concerns to materialize thus far, competition researchers worried about future data scarcity for training FM models have proposed a range of solutions that could do more harm than good. For instance, mandates on digital platforms to share their data

---

world; Comment of US Federal Trade Commission to the US Copyright Office, Artificial Intelligence and Copyright (Oct 30, 2023), 4.

[9] Pablo Villalobos et al., "Will We Run out of Data? An Analysis of the Limits of Scaling Datasets in Machine Learning," *arXiv* preprint, arXiv:2211.04325 (2022).

[10] Human user/programmer feedback, such as ranking outputs generated by a model, are used to teach the model or applications built upon it to better predict how to respond to queries. This is known as reinforcement-based learning. See, Long " et al., "Training Language Models to Follow Instructions with Human Feedback," NEURIPS (2022), https://proceedings.neurips.cc/paper_files/paper/2022/file/b1efde53be364a73914f58805a001731-PaperConference.pdf.

[11] Deepa Seetharaman, "For Data-Guzzling AI Companies, the Internet Is Too Small," *Wall Street Journal*, April 1, 2024, https://www.wsj.com/tech/ai/ai–training–data–synthetic–openai–anthropic–9230f8d8.

[12] Ilia Shumailov et al., "AI Models Collapse When Trained on Recursively Generated Data," *Nature* 631 (2024), https://www.nature.com/articles/s41586-024-07566-y.

[13] Federal Trade Commission, Partnerships Between Cloud Service Providers and AI Developers: FTC Staff Report on AI Partnerships & Investments 6(b) Study, 12–14.

[14] See: Comment of the US Federal Trade Commission to the US Copyright Office, Artificial Intelligence and Copyright, Docket No. 2023-6 (Oct. 30, 2023), 4, https://www.ftc.gov/legal-library/browse/advocacy-filings/comment-federal-trade-commission-artificial-intelligence-copyright.

with rivals and bright-line prohibitions on mergers or business practices that may give firms more control over data, but that may otherwise benefit competition and innovation. These proposed fixes would impose costs on the creators and developers of large digital platforms that FM developers rely upon for deploying and refining their models, and for licensing data used to train the models. They could also reduce competition by deterring the creation and improvement of both the platforms themselves as well as FMs built by their owners. These proposals are thus unlikely to be warranted as ex post antitrust remedies for exclusive dealing or monopolization are adequate for addressing potential abuses of market power. There are also regulatory proposals similar to the European Union's Global Data Protection Regulation (GDPR) that would treat individual digital platform user data as a user's property right. These too are flawed and would worsen data scarcity issues and digital platform functionality while erecting further cost barriers to competition in both the digital platform and generative AI development space. Current legal frameworks and remedies recognized by antitrust courts provide an appropriate, flexible and adequate framework for addressing data monopolization issues should they arise.

This paper will critically appraise current and potential future concerns about data scarcity and data monopolization in generative AI markets and whether these concerns warrant antitrust intervention. It will also argue for the efficacy of current antitrust law remedies in addressing these concerns. To this end, Section I will analyze what counts as a "barrier to entry," and whether that term applies to data in the generative AI space under current or foreseeable future market conditions. Section II will evaluate the likelihood of increasing demand for training data in generative AI innovation in relation to evidence of emerging trends, techniques, and innovations that are likely to reduce data demands and increase the cost efficiency of smaller datasets. The use of synthetic data and other engineering workarounds as well as the trend towards smaller datasets for more targeted AI applications are included in this discussion. Section III will examine how current popular regulatory proposals aimed at addressing the perceived role of data as a barrier to entry could potentially harm competition and innovation. Section IV will analyze whether existing antitrust laws and remedies can sufficiently address future barrier of entry issues around data or the potential abuse of market power over data.

Section I: Data as a Barrier to Entry in the Present

Upfront entry costs exist in virtually every business and industry. If potential market entrants choose not to compete with established incumbents due to low expectations of success or long-term profitability, this outcome could be optimal for consumer welfare and competition. In such cases, the market may already be functioning efficiently with the existing number of firms rather than the limited number of firms or market structure being an indicator of inefficiency or barriers to entry.[15] The potential entrant's scarce resources would be more profitably invested elsewhere, such as in other markets or business lines. For instance, in a market consisting of only a few firms that benefit from scale economies, the potential reduction in output from limited competition may be offset by the increased output resulting from the cost efficiencies of the incumbent firms.[16] There would thus be no benefits to consumers from additional market entry. Rather, there may be a loss of consumer welfare if external interventions lower entry costs, as this could lead to misallocation of resources away from where they could be more productively used.

---

[15] See: C.C. Von Weizsacker, "A Welfare Analysis of Barriers to Entry," *Bell Journal of Economics* 11, no. 2 (Autumn 1980): 399–420.

[16] Ibid.

Some researchers describe the advantages that incumbent firms enjoy in generative AI due to data network effects as "a sizeable entry barrier, particularly for small and medium-sized (SME) enterprises."[17] However, it is unclear whether an incumbent's ready access to data that would-be challengers lack is necessarily a "barrier to entry" in the sense that warrants antitrust intervention.

The price of acquiring data for training generative AI models is a capital cost. Competition policy researchers Dirk Auer and Geoffrey Manne note that "[b]ecause data is used to improve the quality of products and/or to subsidize their use, the idea of data as an entry barrier suggests that any product improvement or price reduction made by an incumbent could be a problematic entry barrier to any new entrant. This is tantamount to an argument that *competition itself* is a cognizable barrier to entry."[18] From an antitrust standpoint, an entry cost should be considered a barrier to entry only if it delays or hinders market entry in a way that reduces consumer surplus or welfare compared to the potential effects of any policy, regulatory, or legal intervention intended to address it.[19] Where capital markets function efficiently and without high levels of uncertainty, scale economies and capital requirements (including data) are not a barrier to entry for potential market entrants regardless of size.[20]

There are several ways that new entrants can obtain data or the means of constructing datasets when entering generative AI markets. New entrants can access capital funding through private and public markets, and independent LLM developers raise hundreds of millions of dollars or even multibillion dollars through institutional investors and venture capitalists.[21] Developers can purchase, license, or construct training datasets through web scraping or through a thriving market of open sources as well as closed (proprietary) sources.[22] Secondary data providers, including data brokers, data warehouses, and synthetic data providers, have entered and continue entering the market to meet growing demands.[23] Rightsholders of copyrighted works can also license these works to AI developers for training or fine-tuning models.[24] Digital platforms like

---

[17] Darek Haftor et al., "A Pathway to Bypassing Market Entry Barriers from Data Network Effects: A Case Study of a Start-Up's Use of Machine Learning," *Journal of Business Research* 168 (2023): 114244.

[18] Dirk Auer and Geoffrey Manne, "From Data Myths to Data Reality: What Generative AI Can Tell Us About Competition Policy (and Vice Versa)," *Competition Policy International* (Feb 2024), https://laweconcenter.org/resources/from-data-myths-to-data-reality-what-generative-ai-can-tell-us-about-competition-policy-and-vice-versa/.

[19] See: McAfee et al., "What is a Barrier to Entry?"

[20] Ibid.

[21] Prominent examples as of March 2024 include Aleph Alpha ($641 million), Cohere ($435 million), Anthropic ($5 billion), AI21 Labs ($321 million), and Mistral AI ($553 million). See: Jonathan Barnett, "The Case Against Preemptive Antitrust in the Generative Artificial Intelligence Ecosystem," in *Artificial Intelligence and Competition Policy*, eds. A. Abbott and T. Schrepel, *Concurrences* (2024), USC CLASS Research Paper 2419 (2024), 665.

[22] See: UK Competition and Markets Authority, "AI Foundation Models: Initial Report," September 18, 2023; Barnett, "The Case Against Preemptive Antitrust in the Generative Artificial Intelligence Ecosystem," 666.

[23] Ibid. See also: Katie Paul, "AI Dataset Licensing Companies Form Trade Group," Reuters, June 26, 2024, https://www.reuters.com/technology/artificial-intelligence/ai-dataset-licensing-companies-form-trade-group-2024-06-26/; Synthesis AI, "Synthesis AI Launches Enterprise Synthetic Dataset on Snowflake Marketplace," PR Newswire, May 11, 2023, https://www.prnewswire.com/news-releases/synthesis-ai-launches-enterprise-synthetic-dataset-on-snowflake-marketplace-301821525.html .

[24] Federal Trade Commission, Partnerships Between Cloud Service Providers and AI Developers: FTC Staff Report on AI Partnerships & Investments 6(b) Study, (February, 2025), 12, 26, https://www.ftc.gov/system/files/ftc_gov/pdf/p246201_aipartnerships6breport_redacted_0.pdf.

Meta have trained their AI models partially on public posts from users of their services, including Instagram and Facebook.[25] These public posts can also be scraped by anyone.

Commensurate with the availability of training datasets and despite concerns about potential data scarcity, the market for generative AI foundation models as of today remains fiercely competitive and dynamic.[26] Current trends indicate an increase in competition and market entry rather than a decline. In September 2023, OpenAI's GPT-4 (which powers ChatGPT), held the vast majority of the market share and had a commanding technological lead over alternatives from big tech platforms like Google and Meta.[27] By November 2024, 16 different AI labs had produced models that exceed GPT-4's capabilities and performance benchmarks.[28] Though some of these models, including Google's Gemini and Meta's LLaMA, have been produced by established tech giants, OpenAI's GPT remains the market leader based on usage. As of January 2025, ChatGPT (59.5 percent) and Microsoft's Copilot (14.3 percent), both of which were built off the GPT LLM, collectively hold a 73.8 percent market share among generative AI chatbot tools.[29] Google's Gemini is in third place with a 13.5 percent market share. In terms of quarterly user growth, however, ChatGPT falls into fourth place behind ClaudeAI from start-up Anthropic, Perplexity (which was built off Meta's LLaMA LLM), and Phind (a tool for software developers, which was built off both GPT and original LLMs). LLM start-ups continue to successfully enter the market, even outside of the text generation segment. For instance, RunwayML has thrived in video generation, and MidJourney and Stable Diffusion thrive in image generation.[30]

Continued and successful market entry by a range of firms of varying sizes that compete head-to-head or serve different generative AI market segments indicates a fiercely competitive and rapidly evolving generative AI and LLM market. Notably, generative AI markets feature a mix of independent start-ups, start-ups benefiting from commercial partnerships with larger entities (such as Amazon/Anthropic and Microsoft/OpenAI), and large tech firms. This contradicts the claim that digital platforms with large troves of user data accumulated through network effects and scale economies possess an unassailable competitive advantage. Simultaneously, partnerships with large platforms have benefited start-ups by providing access to the platform's user base, scale economies, and network effects, all of which help to further develop and refine products. These partnerships also foster the exchange of capital investment, cloud storage, access to computing power, and other resources.[31] Evidently, superior algorithms

---

[25] Mike Clark, "Privacy Matters: Meta's Generative AI Features, META, Sept. 27, 2023, https://about.fb.com/news/2023/09/privacy-matters-metas-g"enerative-ai-features/.

[26] See: Anton Korinek and Jai Vipra, "Concentrating Intelligence: Scaling and Market Structure in Artificial Intelligence," *Economic Policy* 40.121 (2025): 225–56.

[27] Anton Korinek and Jai Vipra, "Market Concentration Implications of Foundation Models: The Invisible Hand of ChatGPT" (Brookings Center on Regulation and Markets Working Paper #9, 2024), https://www.brookings.edu/articles/market-concentration-implications-of-foundation-models-the-invisible-hand-of-chatgpt/.

[28] Korinek and Vipra, "Concentrating Intelligence."

[29] FirstPageSage, "Top Generative AI Chatbots by Market Share – January 2025," Jan 21, 2025, https://firstpagesage.com/reports/top-generative-ai-chatbots/ .

[30] Auer and Manne, "From Data Myths to Data Reality"; Barnett, "The Case Against Preemptive Antitrust in the Generative Artificial Intelligence Ecosystem," 666.

[31] For instance, Microsoft's partnership with OpenAI provided the latter with $14 billion in investment, see: Ananya Gairola, "Microsoft Invested Nearly $14 Billion In OpenAI But Now Its Reducing Its Dependence on the ChatGPT-Parent: Report," Yahoo! Finance, December 25, 2024, https://finance.yahoo.com/news/microsoft-invested-nearly-14-billion-000512844.html. Similarly, Amazon's partnership with Anthropic has given the latter access to billions in investment, as well as resources including Amazon's chips, cloud storage, and the Amazon Web Services platform. Microsoft, Google and Amazon all offer

and preferred user experiences remain a far more significant competitive advantage than access to data, further indicating that data access is not a barrier to entry.[32] The top four LLMs ranked by performance benchmarks as of November 2024 are all either newly launched or have been updated within the preceding three months, which indicates that training data cost or scarcity has not been a barrier to entry or innovation.[33]

The range of firms of different sizes that compete with one-another, including those born out of commercial partnerships between firms of different sizes, also contradicts the theory that large tech giants raise anticompetitive concerns when they move into adjacent AI-related lines of business and use advantages from their existing business lines, including access to platform data. Some fear that tech giants' advantages allow them to restrict competition, and hold an unassailable advantage.[34] Instead, the opposite is observed. The rise of OpenAI's ChatGPT has driven competition and innovation in the adjacent search engine market. Although Google remains the dominant market leader,[35] it is in the face of increasing competition. This is despite Google's monopoly over search indexing and user data across billions of customers on its many mobile and web applications and tools.[36] Thirteen million American adults used generative AI as their primary tool for online searches in 2023, a figure expected to exceed 90 million by 2027.[37] In the face of this threat, Google and Bing responded to ChatGPT's rise in 2023 by integrating generative AI tools into their search engines.[38] OpenAI expanded into search tools by launching its GPT FM-powered SearchGPT tool in October 2024.[39] Since then, it has grown exponentially at a rate that far outstrips the growth of Google's Gemini AI-powered search tool.[40] This demonstrates that generative AI startups are not only competitive against large tech giants, they

---

cloud credits to AI startups, see: Hayden Field and Kif Leswing, "Generative AI 'FOMO' Is Driving Tech Heavyweights to Invest Billions of Dollars in Startups," CNBC, March 30, 2024, https://www.cnbc.com/2024/03/30/fomo-drives-tech-heavyweights-to-invest-billions-in-generative-ai-.html.

[32] It is noted that ChatGPT also benefits from a first-mover advantage, such as through widespread brand recognition and familiarity, which even relatively new tools from tech giants lack. However, the role of this as a competitive advantage in isolation is likely to diminish over time as demonstrated by significantly higher rates of user growth among its newer competitors.

[33] Specifically, OpenAI's GPT, Google's Gemini, xAI's Grok and Anthrophic's Claude, see: Korinek and Vipra, "Concentrating Intelligence."

[34] "[D]ata collected in the origin market can be used, once the enveloper has entered the target market, to provide products more efficiently in the target market … data collected in the origin market can [also] be used to reduce the asymmetric information to which an entrant is typically subject when deciding to invest (for example, in R&D) to enter a new market. For instance, a search engine could be able to predict new trends from consumer searches and therefore face less uncertainty in product design." See: Daniele Condorelli and Jorge Padilla, "Harnessing Platform Envelopment in the Digital World," *Journal of Competition Law and Policy* 16 (2020), 143, 167 cited in Auer and Manne, "From Data Myths to Data Reality."

[35] "566 Million People Used ChatGPT in December 2024 (still behind Google's 6.5 billion but growing fast)," see: Zulekha Nishad, "Is ChatGPT Challenging Google's Dominance?," *Stan Ventures*, Feb 11, 2025, https://www.stanventures.com/news/is-chatgpt-challenging-googles-dominance-1944/ .

[36] Statista and Semrush, "Trends Study: Online Search After ChatGPT," Feb. 2025, https://static.semrush.com/file/docs/evolution-of-online-after-ai/Online_Search_After_ChatGPT.pdf.

[37] Ibid.

[38] Ibid.

[39] Even prior to the launch of SearchGPT, 46 percent of ChatGPT's users were turning on its search feature and using it as a search engine, see: Zulekha Nishad, "Is ChatGPT Challenging Google's Dominance?," *Stan Ventures*, Feb 11, 2025, https://www.stanventures.com/news/is-chatgpt-challenging-googles-dominance-1944/.

[40] Statista and Semrush, "Trends Study: Online Search After ChatGPT," Feb. 2025, https://static.semrush.com/file/docs/evolution-of-online-after-ai/Online_Search_After_ChatGPT.pdf.

are also able to expand into adjacent markets, including those currently dominated by one or more tech giants. This is possible even when startups do not begin with access to the vast user data, network effects, and scale economies of these giants.

Section II: Data as a Barrier to Entry in the Future

It is doubtful that generative AI markets, despite being initially competitive and disruptive, will inevitably converge on just a few players with concentrated and stable market shares. As shown, the history of generative AI markets until today contradicts antitrust regulators' concerns about data scarcity, data network effects, and economies of scale giving an unassailable advantage to tech giants. Concerns about eventual market convergence are instead based on trends observed in mature digital platform markets, such as search and social media, where a few companies with greater user-network effects provide services with high switching costs for users, thereby creating a potentially unassailable advantage. Notably, even these claims about digital markets in general have been repeatedly contradicted in practice.[41] Regardless, the three main features that supposedly make digital platform markets potentially prone to convergence on a few players, including the role of data, do not apply equally to the market for generative AI tools and models.[42] These features are economies of scale, network effects, and switching costs.

First, convergence on a few players is more likely for digital platform markets as they are characterized by economies of scale that advantage incumbents who have already paid high fixed costs and face relatively lower variable costs in the long run, unlike potential new entrants who face high entry costs.[43] Thus, only the most efficient digital platform competitors will remain after others exit the market. Although building a generative AI foundation model does entail high fixed costs, such as pretraining and acquiring data and computing resources, variable costs are nonetheless more significant for market participants than they are for many digital platforms as models must be continuously tested and refined to ensure that they remain fit for their intended purpose and adapt to new information and user demands.[44] The availability of open-source datasets for building new models drastically lowers fixed costs,[45] as do open-source or commercially licensable pre-trained foundation models for building new applications. Large

---

[41] As summed up by Dirk Auer and Geoffrey Manne, "Google overthrew Yahoo, despite initially having access to far fewer users and far less data; Google and Apple overcame Microsoft in the smartphone OS market despite having comparatively tiny ecosystems (at the time) to leverage; and TikTok rose to prominence despite intense competition from incumbents like Instagram, which had much larger user bases. In each of these cases, important product-design decisions (such as the PageRank algorithm, recognizing the specific needs of mobile users, and TikTok's clever algorithm) appear to have played a far greater role than initial user and data endowments (or lack thereof)," see: Auer and Manne, "From Data Myths to Data Reality"; Luxia Le, *The Real Reason Windows Phone Failed Spectacularly*, History–Computer (Aug. 8, 2023), https://history-computer.com/the-real-reason-windows-phone-failed-spectacularly/.

[42] See: Jonathan Barnett, "Illusions of Dominance?: Revisiting the Market Power Assumption in Platform Ecosystems," Revisiting the Market Power Assumption in Platform Ecosystems, USC CLASS Research Paper No. CLASS22-29, USC Law Legal Studies Paper 22–29 (2023); Barnett, "The Case Against Preemptive Antitrust in the Generative Artificial Intelligence Ecosystem," 670.

[43] Ibid.

[44] Christophe Carugati, "The Generative AI Challenges for Competition Authorities," *Intereconomics* 59.1 (2024): 16.

[45] "[P]ublicly accessible and open-source datasets are becoming increasingly available. Notable examples include LAION, a dataset with 5.85 billion image-text pairs generated from CLIP; COYO-700 M, which contains 747 million image-text pairs; the Public Multimodal Dataset, featuring 70 million image-text pairs with 68 million unique images; Common Crawl, which includes data from 50 billion web pages; and The Pile, an 825 GB dataset created by EleutherAI from 22 diverse high-quality subsets," see: Thibault Schrepel and Alex Pentland, "Competition Between AI Foundation Models: Dynamics and Policy Recommendations" (2023), 4.

digital platforms with stable and significant market shares also have an incentive to lower costs for generative AI entrants. For instance, Meta has opted to release the code for its LLaMA foundation models on an open-source basis in order to lower entry costs and attract AI developers to integrate new applications into (and thus improve) its social media platform and "ecosystem".[46] xAI (the AI division of social media platform X) has also released its Grok foundation model on an open source basis.[47] This supports the theory that fixed capital costs do not constitute a barrier to entry for AI models because would-be entrants have equal access to the technology relied upon by incumbents.[48] This is further affirmed by the fact that a majority of popular generative AI foundation models that have entered the market, including OpenAI's GPT-4 and xAI's Grok, were trained entirely on data publicly available on the web.[49]

Second, digital platform markets are more likely to experience convergence on one or a few providers due to user network effects because users seek providers that are popular with other users or whose services are improved by their larger existing user bases.[50] For instance, the quality of Google's search results is refined through data from its large user pool,[51] and Meta is popular as a social media platform due to its volume of existing users. However, users do not attribute value to generative AI models or applications based on the number of other users.[52] Training models on larger datasets of users and their feedback may enhance model quality due to having a broader sample to "learn" from and calibrate responses to. Models that possess more data parameters also benefit from large datasets as more data points allow for the drawing of more inferences about the interrelationships among the model's parameters.[53] However, the role of raw user numbers declines significantly with the degree of specialization in the model or the application's function because improvements in the model or application's output depend more

---

[46] See: Mark Zuckerberg, "Open Source AI Is the Path Forward," Meta, July 23, 2024, https://about.fb.com/news/2024/07/open-source-ai-is-the-path-forward/; Diane Coyle and Hayane Dahmen, "Open Source Generative AI from a Competition Policy Perspective," in *Artificial Intelligence and Competition Policy*, eds. A. Abbott and T. Schrepel, *Concurrences* (2024), 18; Dylan Patel and Afzal Ahmad, "Google 'We Have No Moat, and Neither Does Open AI': Leaked Internal Google Document Claims Open Source AI Will Outcompete Google and OpenAI," *Semianalysis* (May 4, 2023), https://semianalysis.com/2023/05/04/google-we-have-no-moat-and-neither/. Although the code is freely available and can be freely used for research or commercial purposes with some exclusions, a special license must be requested from Meta if the licensee incorporates LLaMA into products or services that amass over 700 million active users. See: Viki Auslender, "Why Meta's Open Source Is Not Really Open," *Ctech by Calcalist*, August 30, 2023, https://www.calcalistech.com/ctechnews/article/atv6xnkya. Pretrained "open source" models may negate or reduce the need for further training data sets, thus lowering entry barriers for developers. However, many "open source" models like X's Grok and Meta's LLaMA models are not transparent about their training data or machine learning processes. This makes it harder for developers to detect or remedy biases in their applications. See: Pascale Davies, "Sorry Elon, Grok Is Not Open Source AI. Here's Why, According to the Creator of the Definition," *Euro News*, March 28, 2024.

[47] Kriti Barua, "What is Grok 3? Elon Musk's 'Smartest AI on Earth' To Be Released Today! Check Details Here," *Jagran Josh*, Feb 17, 2025, https://www.jagranjosh.com/general-knowledge/what-is-grok-3-ai-1739787494-1.

[48] See: McAfee et al., "What Is a Barrier to Entry?"

[49] See: Saffron Huang and Divya Siddarth, "Generative AI and the Digital Commons," *arXiv* preprint arXiv*:2303.11074* (2023); Kriti Barua, "What is Grok 3?"

[50] Barnett, "The Case Against Preemptive Antitrust in the Generative Artificial Intelligence Ecosystem," 670.

[51] "General search services, search advertising, and general search text advertising require complex algorithms that are constantly learning which organic results and ads best respond to user queries; the volume, variety, and velocity of data accelerates the automated learning of search and search advertising algorithms," see: Complaint, United States v. Google, 1:23-cv-00108 (E.D. Va. 2023), https://www.justice.gov/opa/pr/justice-department-sues-google-monopolizing-digital-advertising-technologies.

[52] Barnett, "The Case Against Preemptive Antitrust in the Generative Artificial Intelligence Ecosystem," 671.

[53] Thibault Schrepel and Alex Pentland, "Competition Between AI Foundation Models: Dynamics and Policy Recommendations" (2023), 4.

upon the dataset's quality and suitability for purpose than its size.[54] For instance, an AI tool or model developed for facial recognition in a large American city will produce more accurate results when trained on a dataset of 100,000 racially and gender-diverse faces than when trained on a dataset of 1,000,000 faces of a single gender and/or race. Thus, the appropriateness and scope of a specific training data set for its intended use are more crucial for generative AI development and model training than the data set's size or the number of users from which it may be harvested. This is why a chatbot trained on a larger set containing incomplete or contradictory information—information lacking context or misinformation—will generally produce lower quality and less accurate responses to queries than one trained on a smaller albeit representative data set that lacks acontextual information, contradictory data inputs, or misinformation. Indeed, multiple studies have found that models trained on smaller high-quality datasets can outperform larger models on accuracy and efficiency.[55] In addition to requiring less data due to having fewer data parameters, "smaller language models" and those tailored for specialized tasks also generally have less computational power and memory requirements.[56]

Technological advances, like synthetic data, are also reducing the need for original data and how much it costs to train and fine-tune models. Synthetic data replicates real-world data by making a sample of artificial data that is representative of an existing dataset's features.[57] It can "combat the data scarcity, privacy concerns, and algorithmic biases commonly used in machine learning applications."[58] Although synthetic data comes with the same limitations as the sample of real-world data from which it is derived,[59] and cannot completely replace real-world data, models trained on a combination of synthetic and real-world data have outperformed models trained on larger datasets of real-world data on specific functions.[60] Overreliance on synthetic data for machine learning can result in "model collapse," whereby the accuracy or quality of the model's results decline over time.[61] However, synthetic data techniques and the quality of synthetic data are also likely to be refined over time and with the benefit of technological improvements.

Decentralized AI is another innovation that promises to reduce entry barriers to AI development by making data more accessible. At the time of writing, most major AI models (like OpenAI's GPT, Meta's LLaMA, Google's Gemini, Anthropic's MistralAI, etc.) are built on centralized servers that host datasets and code. They are controlled by centralized corporations

---

[54] UK Competition and Markets Authority, "AI Foundation Models: Initial Report," September 18, 2023, 56–57.

[55] See: Rohan Anil et al., PaLM 2 Technical Report, Computer Science (September 13, 2023) cited in Fausto Gernome and David Teece, "Competing in the Age of AI: Firm Capabilities and Antitrust Considerations," in *Artificial Intelligence and Competition Policy*, eds. A. Abbott and T. Schrepel, *Concurrences* (2024), 73; X. Geng et al., "Koala: A Dialogue Model for Academic Research," *The Berkeley Artificial Intelligence Research Blog*, 2023, https://perma.cc/9HUC-K9KC.

[56] UK Competition and Markets Authority, "AI Foundation Models: Initial Report," September 18, 2023, at 10–11; Barnett, "The Case Against Preemptive Antitrust in the Generative Artificial Intelligence Ecosystem," 668.

[57] Sergey Nikolenko, "Synthetic Data for Deep Learning," *Springer Nature* 174 (2021).

[58] Mandeep Goyal and Qusay H. Mahmoud, "A Systematic Review of Synthetic Data Generation Techniques Using Generative AI," *Electronics* 13, no. 17 (2024): 3509, https://doi.org/10.3390/electronics13173509.

[59] Since synthetic data is based on an existing sample of real-world data, it also replicates any biases or deficiencies in the sample or any features that make it less representative of what it is meant to be a sample of, as well as the biases of those who collected the original data, and the biases of the data engineers feeding the model contextual information.

[60] For example, MIT researchers used a combination of real-world and synthetic images to train an image generation model, with the result being that it outperformed competitor models in accuracy even though those models were trained on a much larger dataset of real-world images. See: Rachel Gordon, "Synthetic Imagery Sets New Bar In AI Training Efficiency," MIT News, November 20, 2023, https://news.mit.edu/2023/synthetic-imagery-sets-new-bar-ai-training-efficiency-1120.

[61] Shumailov et al, "AI Models Collapse When Trained on Recursively Generated Data."

that hire workers for machine learning, data engineering, data cleaning, and coding. These corporations, or those they enlist, are also responsible for acquiring data, computation power (hardware), cloud storage, energy, and other necessary inputs. By contrast, decentralized AI involves using blockchain technology to host AI models and their underlying datasets and other components across a decentralized, collaborative network of servers (or "nodes"), thereby allowing for open, real-time, and continuous collaboration between individuals.[62] Digital assets, such as cryptocurrency tokens, are provided as rewards to those who contribute to the network by providing data, computing power, coding, machine learning, data cleaning, application integration, or hosting services that support the project.[63] "Subnets" are another way that decentralized AI can foster competition and counter potential data entry barriers to generative AI development in the future. These are smaller networks of collaborators within the larger decentralized blockchain network that focus on specific projects. For instance, Bittensor is an open-source decentralized AI protocol maintained and developed by the Opentensor Foundation.[64] The protocol is controlled and governed through community decisions across servers on its blockchain and any individual can contribute, with the foundation providing light oversight.[65] It includes 64 active subnets as of February 2025.[66] These include "Subnet 42 Real-Time Data," which was launched by AI startup Masa, and provides data access with the goal of "democratizing" AI development.[67] These and other innovations are likely to continue reducing potential competitive barriers that could result in the future from data scarcity.

Besides technological innovation, human creativity and ingenuity also foster new means for lowering machine learning and data costs. For instance, the Chinese open-source AI foundation model DeepSeek delivers a comparable performance to leading models such as OpenAI's GPT despite being trained at a significantly lower cost.[68] DeepSeek's success in lowering training dataset and computation demands has been attributed to "a combination of many smart engineering choices including using fewer bits to represent model weights, innovation in the neural network architecture, and reducing communication overhead as data is passed around between GPUs."[69] The public release of DeepSeek's code and model weights means that these techniques can be replicated by other developers, signaling that further creative fixes in this space that lower costs and reduce entry barriers are possible.

Even when a large, high-quality dataset exists for model training or application refinement, the value of adding more data decreases after a certain point. A facial recognition model trained on a population-representative sample of 1,000,000 faces may not witness significant improvements in its performance with additional inputs of faces, and additional faces

---

[62] See: Justin Harris and Bo Waggoner, "Decentralized and collaborative AI on blockchain," 2019 IEEE International Conference on Blockchain (Blockchain), IEEE, 2019.

[63] Ibid.

[64] Tor Constantino, "How Big Tech, ChatGPT and DeepSeek Could Lose To Decentralized AI," *Forbes*, Feb 12, 2025.

[65] Ibid.

[66] Ibid.

[67] Ibid.

[68] Graham Fraser, "DeepSeek vs ChatGPT - How Do They Compare?" BBC.com, 28 January, 2025, https://www.bbc.com/news/articles/cqx9zn27700o.

[69] Ambuj Tewari, "Why Building Big Ais Costs Billions – and How Chinese Startup Deepseek Dramatically Changed the Calculus," *The Conversation*, January 29, 2025, https://theconversation.com/why-building-big-ais-costs-billions-and-how-chinese-startup-deepseek-dramatically-changed-the-calculus-248431#. GPUs or Graphics Processing Units are electronic circuits that produce complex calculations at high speed.

would not allow it to attract customers from a competitor application trained on its existing dataset that has a better algorithm, provides a better user experience, or is more cost or energy efficient to run. Even when additional data inputs would significantly improve model performance, this may still not be enough to outweigh a competitor's improved algorithm, user interface, or cost efficiency. If two generative AI competitors have enough of the right kind of data, improvements in these other areas would give a greater competitive advantage in determining competitive outcomes. This is supported by studies that find that even for digital platforms, data network effects and additional data have diminishing marginal returns.[70] Furthermore, the cost of data cleaning and categorization (a necessary part of machine learning and a process that improves model and application performance given the same dataset) increases with the size of a dataset.[71] Beyond a certain point, the performance gains from increasing the dataset size may be outweighed by the rising costs of cleaning, categorizing, and processing the additional data.[72]

Finally, digital markets are sometimes prone to convergence on a few players due to the costs incurred by users when they attempt to switch to new or existing competitors.[73] In the case of foundation models, developers could incur costs by having to migrate to alternative foundation models. However, tools and platforms already exist that mitigate these costs by allowing for foundation model-agnostic application development.[74] For instance, LLM-agnostic software architecture, a framework adopted by some companies, allows for generative AI prompts to flow to different FMs depending on the application and its specific requirements. To be included, applications must be coded in a way that divorces their underlying logic from that of any specific or single FM.[75] Model switching is also facilitated by large open source platforms, such as the Hugging Face website, which hosts over 300,000 models and 250,000 datasets as of August 2024.[76] Given these factors and ongoing developments, it is unlikely that access to data will become an antitrust barrier to entry in the foreseeable future.

---

[70] See: Catherine Tucker, "Digital Data, Platforms and the Usual [Antitrust] Suspects: Network Effects, Switching Costs, Essential Facility," *Rev. Indus. Org.* 54 (2019): 683, 686; Auer and Manne, "From Data Myths to Data Reality"; Marco Iansiti, The Value of Data and Its Impact on Competition (Harvard Business School, NOM Unit Working Paper No. 22-002, July 20, 2021).

[71] See: Ryan Aminollahi, "Generative AI for Enhanced Data Cleansing and Management" *Medium*, November 1, 2024, https://ryanaminollahi.medium.com/generative-ai-for-enhanced-data-cleansing-and-management-d8eb7c560a1e#.

[72] Due to their larger sample size and high dimensionality, using larger datasets "introduce[s] unique computational and statistical challenges, including scalability and storage bottleneck, noise accumulation, spurious correlation, incidental endogeneity and measurement errors," see: J. Fan et al., "Challenges of Big Data Analysis," *National Science Review* (2014) 1, 293–314.

[73] Jonathan Barnett, "Illusions of Dominance?: Revisiting the Market Power Assumption in Platform Ecosystems," Revisiting the Market Power Assumption in Platform Ecosystems (May 18, 2023). USC CLASS Research Paper No. CLASS22-29, USC Law Legal Studies Paper 22-29 (2023); Barnett, "The Case Against Preemptive Antitrust in the Generative Artificial Intelligence Ecosystem," 670.

[74] See: Barnett, "The Case Against Preemptive Antitrust in the Generative Artificial Intelligence Ecosystem," 273; Adrian Bridgewater, "It's Time To Believe In AI Agnosticism," *Forbes*, February 12, 2024, https://www.forbes.com/sites/adrianbridgwater/2024/02/12/why-its-time-to-believe-in-ai-agnosticism/.

[75] See e.g. *Entrio*, "Implementing an LLM Agnostic Architecture for our Generative AI Module," March 20, 2024, https://www.entrio.io/blog/implementing-llm-agnostic-architecture-generative-ai-module.

[76] Jonathan Gillham, HuggingFace Statistics, Origniality.ai (August 8, 2024), https://originality.ai/blog/huggingface-statistics.

Section III: Analyzing Proposed Regulatory Fixes: Bright-Line Rules and Data Rights

Some policymakers and regulators believe that a proactive regulatory approach—imposing controls and guardrails—is necessary to ensure society benefits from AI technology while mitigating risks such as data misuse and anticompetitive concerns related to data control.[77] However, such regulations should only be imposed if their benefits to competition and consumers exceed their costs (inadvertent or otherwise) relative to a world where AI technology is allowed to develop without the proposed regulation.[78] Regulations should only be imposed upon markets where there is a "clear market failure, [and] also strong grounds for believing that government intervention will yield net welfare-superior results to non-intervention."[79] Potential harms of regulation, including those introduced in the name of promoting competition, include regulatory capture that benefits vested interests while raising rather than reducing entry barriers for new or less politically and economically-resourced entrants.[80] Regulations also typically impose significant fixed costs, which place a disproportionate burden on smaller competitors in an industry or market relative to their larger peers as larger firms experiencing scale economies can spread the costs of compliance.[81]

AI is a rapidly evolving technology and generative AI markets are characterized by rapid growth.[82] Many of the markets around AI technology and its resulting products have yet to form, making the potential adverse consequences of overregulation hard to predict. Since AI innovations will power productivity in virtually every major industry and market, the adverse economic consequences of overregulation could be especially dire.[83] There could be a tremendous

---

[77] See, e.g. Francesco Filippucci et al., "The Impact of Artificial Intelligence on Productivity, Distribution and Growth: Key Mechanisms, Initial Evidence and Policy Challenges," OECD Artificial Intelligence Papers (April 2024) cited in John M. Yun, "The Folly of AI Regulation," in *Artificial Intelligence and Competition Policy*, eds. A. Abbott and T. Schrepel, *Concurrences* (2024), 615; The White House, Executive Order on the Safe, Secure and Trustworthy Development and Use of Artificial Intelligence, (October 30, 2023), https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence. It is further noted that a firm's ability to misuse consumer data could be a sign that it possesses market power. See: Howard A. Shelanski, "Information, Innovation, and Competition Policy for the Internet," *University of Pennsylvania Law Review* 161 (2013): 1663.

[78] See: Thomas Lambert, *How to Regulate: A Guide for Policymakers* (Cambridge University Press, 2017); Harold Demsetz, "Information and Efficiency: Another Viewpoint," *Journal of Law and Economics* 12, no. 1 (1969).

[79] Yun, "The Folly of AI Regulation," 618.

[80] "[R]egulations can have the consequence of entrenching and raising barriers to entry, which, perversely, harms the competitive process rather than promoting it." See: Yun, "The Folly of AI Regulation," 621. See also: Thibault Schrepel, "Decoding the AI Act: A Critical Guide for Competition Experts," (Amsterdam Law and Technology Institute, Working Paper No. 3-2023, 2023).

[81] "Regulation has high fixed costs but low marginal costs. The larger a firm is, the more it can spread those fixed costs and minimize their impact on price per unit. One of the unintended consequences of government regulation is imposes relatively greater costs on smaller producers, especially new entrants, and that it rewards scale. So the more you regulate an industry, the more consolidation you'll get." See: Michael F. Cannon, "What's Driving Provider Consolidation?" Cato Institute, May 5, 2021, https://www.cato.org/blog/whats-driving-provider-consolidation. "Ironically, big tech companies such as Facebook, Amazon, Apple and Google benefit from a silver lining when it comes to being regulated — what hurts their competitors more only makes them stronger." See: Jedidiah Yueh, "GDPR Will Make Big Tech Even Bigger," *Forbes*, June 26, 2018, https://www.forbes.com/sites/forbestechcouncil/2018/06/26/gdpr-will-make-big-tech-even-bigger/ .

[82] See: Xuli Tang et al., "The Pace of Artificial Intelligence Innovations: Speed, Talent and Trial-and-Error," *Journal of Infometrics* 14 (November 2020).

[83] Economic research finds that in the long run, all growth in output per worker can be attributed to technological progress. See: Peter Howitt and Philippe Aghion, "Capital, Innovation and Growth Accounting," *Oxford Review of Economics* 23 (2007): 79, 80. "As global industries evolve, the rapid advancement of artificial intelligence (AI) is emerging as a catalyst for transformative change across nearly every sector. While the potential impact of AI spans virtually all industries, three sectors—tax and accounting, healthcare, and transportation—stand out as arenas where AI's influence is both profound and immediate." See: Brent

net loss in societal and consumer welfare from a regulation that retards the emerging technology's exponential growth rate, even if the regulation is successful in reducing certain harms to some degree.[84]

A better approach is "permissionless innovation," whereby new innovations aren't assumed to cause theoretical harm unless the innovator can prove otherwise beforehand.[85] This approach allows new and rapidly evolving technologies to develop while surgically addressing harms as they emerge. For instance, the relatively light-touch approach taken by Congress in the 1996 Telecommunications Act, aimed at addressing concerns about the internet's unfettered development, has been credited with enabling "the tremendous innovation and growth that we are still experiencing today."[86] Preemptive regulation, on the other hand, crowds out market responses that tend to be more efficient and do not cost taxpayers the exorbitant administrative costs of enforcement.[87] When regulations do end up being adopted, they are more likely to be effective while reducing inadvertent damage to competition and innovation where the harms they target are clearly defined rather than vague.

The European Union's [EU's] attempts to regulate the use of data and reduce privacy harms in digital markets demonstrate the adverse impacts of imposing proscriptive regulations on innovative sectors. In 2018, the EU instituted the General Data Protection Regulation (GDPR), which requires companies and organizations to guarantee users the right to access and erase data. The GDPR also mandates positive opt-out style consent for certain data collection and use and that users can request their data and move it between different platforms, that is, render it "portable."[88] When the GDPR was instituted, users did begin opting out of tracking and data collection, which led to less data collection and use by businesses, which hampered competition and innovation among online businesses. Web visits and revenue declined, while online search quality suffered and became less efficient due to reduced access to data, which limited the ability to tailor results to user preferences.[89] As competition in web advertising and user preference tracking decreased, the number of new firms, innovative applications, and venture capital investment also decreased.[90] Difficulties in acquiring data or using it for targeted services like advertising due to the GDPR disproportionately affected small businesses[91] and newer or potential

---

Gleeson, "How AI is Reshaping the Future of Work Across Industries," *Forbes*, December 3, 2024, https://www.forbes.com/sites/brentgleeson/2024/12/03/how-ai-is-reshaping-the-future-of-work-across-industries/.

[84] Yun, "The Folly of AI Regulation," 621.

[85] See: Adam Thierer, Permissionless Innovation: The Continuing Case for Comprehensive Technological Freedom (2016).

[86] Yun, "The Folly of AI Regulation," 620.

[87] Yun, "The Folly of AI Regulation," 621, 630–31. As new purported harms or defects emerge in technological products, businesses have a commercial incentive to remedy these to protect their brand reputation and goodwill (which is easily damaged and which firms sometimes invest billions of dollars into building) and to ensure repeat business. See: Benjamin Klein and Keith Leffler, "The Role of Market Forces in Assuring Contractual Performance," *Journal of Political Economy* 89, no. 4 (1981). An example in the generative AI space is when Google rapidly responded to criticism that its Gemini AI generator was producing historically inaccurate depictions of historical figures with their race or sexuality altered by withdrawing and relaunching the software after fixing the defects. See: William Gavin, "Google's Gemini AI Generator Will Relaunch in a Few Weeks After Spitting Out Inaccurate Images," *Quartz*, Feb 26, 2024.

[88] Li et al., "The impact of GDPR on global technology development," *Journal of Global Information Technology Management* 22.1 (2019): 1–4.

[89] Varina Que and Avi Goldfarb, "The Economics of Digital Privacy," *Annual Review of Economics* 15: 267, 280.

[90] Ibid.

[91] John M. Yun, "A report card on the impact of Europe's privacy regulation (GDPR) on digital markets," *George Mason Law Review* 31, (2024): 104–24.

entrants that lacked an established brand presence predating the GDPR. Because of the GDPR, the productivity of EU firms reduced relative to their American counterparts, and European tech start-ups saw reduced access to investment for research and development, further hampering innovation.[92]

The GDPR's regulatory and compliance costs have also been significant,[93] with firms' privacy office budgets increasing by an average of 29 percent to meet compliance requirements.[94] Additionally, American companies adhering to GDPR and similar privacy laws, such as the California Consumer Protection Act (CCPA), incur an estimated cost of $480 per individual whose data they retain.[95] These costs are more easily borne by tech giants and the firms they own: Google, Amazon, and Meta find it easier to obtain consent from individual users.[96] The costs and their relative burdens have further driven increases in the market concentration of tech giants and the largest firms at the expense of smaller players.[97] The long-term impact of the GDPR on innovation has yet to be fully measured due to the dynamic nature of the affected markets.

The GDPR is also likely to adversely impact European AI development going forward. For instance, Italian regulators banned ChatGPT in 2023 after determining that it had no basis to collect and analyze user data under the GDPR.[98] This prohibits ChatGPT (and potentially other AI firms) from obtaining data from Italian users for fine-tuning (or refining) applications. Similar regulations that impose restrictions on data use in the name of regulating AI will also likely adversely impact AI quality and innovation[99] and harm competition against entrenched incumbents by disproportionately burdening newer and smaller rivals.

---

[92] Jian Jia et al., "The Short-Run Effects of GDPR on Technology Venture Investment (National Bureau of Economic Research., Working Paper No. 25248, 2018); European Commission, Digital Economy and Society Index (Desi) 2021, 6 and 9.

[93] John M. Yun, "A Report Card on the Impact of Europe's Privacy Regulation (GDPR) on Digital Markets"; "For example, according to a 2017 PwC survey more than 40 percent of responding firms spent over $10 million on GDPR compliance efforts. [In 2018, an] EY and International Association of Privacy Professionals report found companies reported spending an average of $1.3 million per year on GDPR compliance costs. These costs are undertaken not only by European companies but also by US-based companies with an EU presence." See: Jennifer Huddleston, "The Price of Privacy: The Impact of Strict Data Regulations on Innovation and More," American Action Forum, June 3, 2021, https://www.americanactionforum.org/insight/the-price-of-privacy-the-impact-of-strict-data-regulations-on-innovation-and-more/.

[94] Müge Fazlioglu, IAPP-EY Annual Privacy Governance Report 2021, Int'l Ass'n of Priv. Pros., xii (2021).

[95] Alan McQuinn and Daniel Castro, "The Costs of an Unnecessarily Stringent Federal Data Privacy Law," Info. Tech. & Innovation Found., Aug. 5, 2019, https://itif.org/publications/2019/08/05/costs-unnecessarily-stringent-federal-data-privacy-law .

[96] See: Garrett A. Johnson et al., "Privacy and Market Concentration: Intended & Unintended Consequences of the GDPR," *Mgmt. Sci.* 69, (2023); J. Campbell., A. Goldfarb, and C. Tucker, "Privacy Regulation and Market Structure," *Journal of Economics & Management Strategy* 24 (2015) (1), 47–73.

[97] Jane Bambauer, "How to Get the Property out of Privacy Law," *Yale Law Journal*, 22 April 2024; Garrett Johnson et al., "Privacy and Market Concentration: Intended & Unintended Consequences of the GDPR."

[98] Karina Tsui, Italy Bans ChatGPT over Privacy Concerns, Semafor (March 31, 2023), https://www.semafor.com/article/03/31/2023/chatgpt-banned-italy-privacy-concerns.

[99] One potential drawback of requiring consent prompts for data use in matching learning is increased difficulty in correcting model biases, which could lead to real-world discrimination by AI tools and applications. Bambauer notes that "[d]ata practices in machine learning or basic social science research depend on having a representative sample of data to perform well and avoid biased results. These goals would be frustrated if some (nonrandom) set of data subjects refuse to allow access to their data…the timely and worthy goals of tackling unintentional bias in AI systems will require *more* personal data to avoid biased and unnecessary error in predictions." [emphasis is in the source]. See: Bambauer, "How to Get the Property out of Privacy Law," 743; Jane Yakowitz Bambauer, "Tragedy of the Data Commons," Harv. J.L. & Tech. 25 (2011): 1, 61, 64 (published as Jane Yakowitz); Sandra G. Mayson, "Bias in, Bias out," *Yale L.J.* 128 (2019): 2218, 2224; Alice Xiang, "Being 'Seen' Versus 'Mis-Seen': Tensions Between Privacy and Fairness in Computer Vision," *Harv. J.L. & Tech.* 36 (2022): 45-49.

In addition to harming competition, regulations intended to protect or otherwise benefit consumers (such as by improving privacy standards and control over data) can also harm them directly. As summarized by University of Florida law professor Jane Bambauer, "[c]ompanies [bearing regulatory compliance costs] respond by raising prices for consumers—a cost that may well be worth it to some people and in some contexts, but probably not as a general rule. And this ignores the costs of inconvenience not only in the form of the time required to click through and manage consents but also in terms of the degraded service that results from a less customized experience. For example, the introduction of GDPR seems to have resulted in consumers having to use 21 percent more search terms and access 16 percent more websites before making their online transactions."[100] So called "protective" regulations on AI tools are thus likely to degrade user experience and product quality while raising the costs for consumers accessing paid services.

Additionally, there must be a cost-benefit calculation for regulations that intend to counter the convergence of the generative AI sector on a few foundation model developers. It may be the case that such convergence and high levels of concentration may come to be the most efficient market structure for competition over time, as efficiencies and technical improvements beyond a certain point may be possible only where a firm can take advantage of network effects and scale economies that arise from concentration.[101] For instance, if a single firm develops a superior foundation model that outcompetes all others by attracting a large and exclusive user base on a single platform, that firm can leverage user interactions to refine and improve its generative AI model. In this situation, breaking up the firm to create more competitors and offer consumers "more options" could backfire by creating multiple inferior products. Additionally, such an intervention could discourage both the firm's successors and potential competitors from investing in developing a superior product, fearing that achieving market dominance would again trigger regulatory intervention.

Since at least the 1966 *Grinnell* case, the US Supreme Court has recognized that market concentration isn't an unqualified bad and can instead be a sign and outcome of vigorous competition rather than an indicator of consumer harm or lack of competition.[102] The same holds true when a firm is able to develop a superior product due to merging with or acquiring another firm or its assets, including data.[103] For a merger or acquisition to be deemed illegal, there must be a substantial likelihood that it will reduce competition in the relevant market.[104] This

---

[100] Jane Bambauer, "How to Get the Property out of Privacy Law," 1103, citing Yu Zhao et al., "Privacy Regulations and Online Search Friction: Evidence from GDPR," *Nat'l Bureau Econ. Rsch.* 1 (Aug. 2021).

[101] See, Barnett, "The Case Against Preemptive Antitrust in the Generative Artificial Intelligence Ecosystem," 672; Barnett, "Illusions of Dominance: Revisiting the Market Power Assumption in Platform Markets."

[102] The Supreme Court ruled that acquisition of monopoly power (such as a level of market share or concentration commensurate with monopoly power) is not illegal anticompetitive behavior in violation of Section 2 of the Sherman Act if it was obtained due to "growth or development as a consequence of a superior product, business acumen, or historic accident," see: United States v. Grinnell Corp., 384 U.S. 563 (1966) *('Grinnell')*.

[103] "[E]conomic efficiencies produced by the merger must be weighed against anticompetitive consequences in the final determination whether the net effect on competition is substantially adverse [thus rendering the merger illegal]." See: FTC v. Procter & Gamble Co., 386 U.S. 568, (1967).

[104] Current merger law prevents the consideration of procompetitive efficiencies in markets outside of the immediate one where the merging firms are deemed to compete (in the case of a horizontal merger.) This incentivizes competition officers to argue for market definitions that limit the scope of the market in order to increase their chances of blocking the merger by excluding "out of market efficiencies." However, if the outcome of this is to prevent a merger that would have benefited consumers in adjacent markets, then competition and consumer welfare would be reduced by blocking the merger. See: FTC v. Procter & Gamble Co., 386 U.S. 568 (1967); Mark J. Niefer and Aaron D. Hoag, "Artificial Intelligence, Uncertainty and Merger Review." Artificial Intelligence and Competition Policy (A. Abbott and T. Schrepel eds.), Concurrences (2024) at 252; John M. Yun, "Reevaluating

determination requires showing that the acquisition causes anticompetitive harm when weighed against likely procompetitive benefits.

Concerns about market power stemming from data acquisition in digital markets like generative AI have prompted some scholars to call for the companies that hold data to share this resource with rivals and to allow users to freely move data between competing services.[105] These data sharing or data portability mandates come with both potential benefits and costs for competition. On one hand, allowing developers and other foundation model creators to access a firm's competitively valuable data—either for free or for a lower price—could foster competition by reducing the cost to compete and build comparable models and applications. This access would also encourage firms to compete on factors beyond data acquisition and use, such as enhancing user experience or modifying how data is used. This approach would reduce other potential purported harms of data gathering and scraping from websites, services, and databases, such as privacy and copyright infringement.

Conversely, however, mandated data sharing would also reduce incentives for firms to gather data for developing superior AI models in the first place, since the expected profits from gathering and monetizing the data would be reduced by rivals being able to free-ride off the same training data. Thus, model developers who might otherwise have scraped or purchased the training data for their own FMs from these firms and application developers who would have built applications off models trained on those superior data sets would be left worse off. In other words, there would be no sense in mandating the forced sharing of a resource if the result would be to discourage that resource from being created in the first place. For instance, data sharing and portability mandates would reduce the incentives for digital platforms, such as search engines or social media or e-commerce sites, to invest in developing and improving the platform since they could no longer profit from the exclusive ability to monetize their data through means such as targeted advertising.[106] Platforms where user experience suffers from reduced incentives to invest in improvement would attract fewer users, and the aggregation of a large pool of user data and interactions would thus be further stymied. Rather than helping other firms, this outcome would erode their ability to get the data that they might otherwise have obtained through scraping it or licensing it from the platform owner. To reiterate, bodies of useful data may never come to exist if data sharing is enforced.[107]

Out of Market Efficiencies in Antitrust, 54 Arizona State Law Journal 1262 (2022). Given that generative AI technology impacts a vast range of markets, there could be adverse outcomes for innovation from failing to weigh out-of-market efficiencies in mergers concerning AI.

[105] See: Nathan Newman, "Search, Antitrust, and the Economics of the Control of User Data," *Yale Journal on Regulation* 31 (2014):, 401, 448.

[106] As summed up by the late Justice Scalia of the US Supreme Court, "[f]irms may acquire monopoly power by establishing an infrastructure that renders them uniquely suited to serve their customers. Compelling such firms to share the source of their advantage is in some tension with the underlying purpose of antitrust law, since it may lessen the incentive for the monopolist, the rival, or both to invest in those economically beneficial facilities." See: Verizon Communications Inc. v. Law Offices of Curtis v. Trinko, LLP, 540 U.S. 39, (2004) at 407-408.

[107] "[P]latforms must often go to great lengths in order to create data about their users — data which these same users often do not know about themselves. Under this framing, data is a by-product of firms' activity rather than an input that is necessary for rivals to launch a business …   there is also information that does not even 'exist' in any real sense (or at least is not known) until the mechanism is created to elicit it … users have no particular comparative advantage in the eliciting or interpreting of that information … it may not even be known ex ante, and, even if it is, in many cases it is virtually useless. As a result, the [platform's] mechanisms that elicit and share that information with others who do have a comparative advantage in using it are of great value to users — not only because the information, once processed, may be used by others in ways that ultimately impart value to the user, but also because the very act of eliciting and sharing the information imparts knowledge to the user directly …

Additionally, data sharing and portability mandates risk compromising user privacy and leaving sensitive data vulnerable to cybercriminals. Competitors that gain access to the data may not have the same privacy standards and protections or the same cybersecurity and anti-hacking resources as the service that originally collected the data. Services that attract users by promising better privacy protections and more user control over the data would be unable to guarantee these standards once the data is shared with other firms. Therefore, the overall risks to user privacy and data security from mandated interoperability and data portability could outweigh any benefits to privacy gained from companies shifting their attention away from data collection.

In sum, regulatory, policy, and legal fixes have been proposed in response to concerns about anticompetitive behavior in the generative AI space. These concerns center on a fear of market failures resulting from just a few firms cornering data that give their technologies an unassailable advantage. However, proposed remedies to anticompetitive behavior, such as breaking up firms or platforms or imposing interoperability or data portability mandates on them, would come with substantial costs that may outweigh their benefits. The impact of new rules for AI innovation and development is hard to predict and a negative impact that even slightly reduces the growth rate of technological development can have an exponential negative economic impact on innovation and competition in the long run. Policymakers should thus err against imposing regulations: (1) that are not surgically targeted against identifiable anticompetitive harms; and (2) where any inadvertent harms to competition and innovation from the regulation are not relatively predictable and cannot be minimized.

For business practices and data acquisitions that are not clearly harmful or have both competitive and anticompetitive effects with uncertain outcomes, a "permissionless innovation" approach should be adopted. This means allowing firms to experiment freely unless clear harms can be identified. Where harms are identified, a new rule can be conceived if enforcing existing ones under current law and policy ex-post are found to be lacking and where the benefits of the newly proposed rule are likely to outweigh its costs relative to the status quo.

Section IV: Sufficiency of Existing Antitrust Remedies in Addressing Anticompetitive Data Concerns

As the preceding sections make clear, antitrust enforcement and policy in high-innovation areas like generative AI should focus on evidence of anticompetitive conduct rather than hypothetical theoretical future harms.[108] Regulators should carefully weigh the likely net positive impact of any policy on competition, innovation, and consumers against unregulated outcomes. New regulations

---

[a]n enormous quantity of the data at issue in these policy discussions [does not exist or is useless] until it is made manifest through some activity by which the user interacts with the platform. Thus the value of those activities is not just in the sharing of information with others, but in the creation of information in the first place." See: Dirk Auer et al., "Comments of International Center for Law and Economics: Understanding Competition in Markets Involving Data or Personal or Commercial Information," International Center for Law and Economics (ICLE), January 7, 2019, 19–20, https://laweconcenter.org/wp-content/uploads/2019/07/Understanding-Competition-in-Markets-Involving-Data-or-Personal-or-Commercial-Information-FTC-hearings-ICLE-Comment-7.pdf.

[108] Merger enforcement targeting proposed deals that will likely have anticompetitive implications is an exception. It is impossible to know what anticompetitive outcomes or practices will be enabled beforehand, thereby necessitating a predictive exercise about probable harms. "[A]s the Supreme Court has emphasized, the Clayton Act [which prohibits mergers and acquisitions that may substantially lessen competition or tend to create a monopoly] was passed to halt trends towards concentration in any particular line of commerce in their 'incipiency,' before competitive conditions have already been eroded." See: Mark J. Niefer and Aaron D. Hoag, "Artificial Intelligence, Uncertainty and Merger Review," 219, 220 citing Brown Shoe Co. v. United States, 370 Y,S, 294, 317.

ought only be considered where current ones and laws are insufficient. In this regard, the United States' existing antitrust laws already provide a dynamic and flexible framework for policing anticompetitive harms. These laws apply to issues arising from the use, misuse, or control of data, enabling the outright prohibition and punishment of inherently anticompetitive conduct, as well as the flexible, case-by-case assessment of business conduct and practices that may or may not have net anticompetitive impacts. The Supreme Court has criticized "[l]egal presumptions that rest on formalistic distinctions rather than actual market realities."[109] In dynamic and rapidly evolving fields like AI where the freedom to experiment and innovate is crucial, and where today's "market reality" may not reflect that of tomorrow, outright prohibitions on business conduct based on its form rather than its anticompetitive substance can be especially damaging to innovation and competition.

Similarly, because enforcement actions and antitrust litigation impose substantial costs and uncertainty on the targeted parties, enforcers should carefully choose when to bring cases against business conduct or mergers with ambiguous impacts on competition and innovation. The risk and threat of multi-year, multimillion-dollar potential litigation costs themselves could deter innovators from procompetitive conduct.[110] To be clear, this is not to suggest that enforcement actions should only be brought where the anticompetitive implications of a business practice can be assumed with absolute certainty.[111] That is a rare occurrence,[112] and such an approach would result in an unacceptable number of anticompetitive behaviors and deals. It would also allow firms that genuinely abuse their market power to further entrench their positions. However, the high cost of false positive errors in identifying and blocking allegedly anticompetitive practices and deals (even relative to the costs of permitting false negatives) should be acknowledged in contemplating enforcement actions and policy.[113] No cure should be worse than the disease.

---

[109] Eastman Kodak Co. v. Image Technical Services, Inc., 504 U.S. 451, 466 (1992).

[110] For instance, "[l]itigating against an FTC or DOJ attempt to block a merger (or appealing an injunction secured against a merger) is a costly, multiyear process that often forces parties to abandon mergers even when the mergers could have survived a court challenge." See: Satya Marar, "Artificial Intelligence and Antitrust Law: A Primer" (Mercatus Special Study, Mercatus Center at George Mason University, March 2024), 15, citing Douglas H. Ginsburg and Joshua D. Wright, "Philadelphia National Bank: Bad Economics, Bad Law, Good Riddance," *Antitrust Law Journal* 80, no. 2 (2015): 496.

[111] For instance, and with regard to enforcement against anticompetitive mergers, "as the Clayton Act makes clear, the government is only required to show that a merger *may* [emphasis original] substantially lessen competition or tend to create a monopoly, not that it will certainly do so. Merger analysis is, in short, an inherently *probabilistic* [emphasis original] undertaking, where uncertainty abounds." See: Mark J. Niefer and Aaron D. Hoag, "Artificial Intelligence, Uncertainty and Merger Review," at 220 citing United States v. Philadelphia National Bank, 374 U.S. 321, 362 (1963), and Hospital Corporation of America v. FTC, 807 F.2d 1381, 1389 (7th Circuit 1986). Further, "courts are, quite properly, unwilling to block mergers based on the mere theoretical possibility of competitive harm. The fundamental question is, therefore, when does the risk of competitive harm go from merely theoretical to likely enough to warrant blocking a merger." In markets like AI that are undergoing rapid change or are in flux, predictions of harm become less certain. See: Mark J. Niefer and Aaron D. Hoag, "Artificial Intelligence, Uncertainty and Merger Review," 221.

[112] Merger enforcement, by its very nature, requires courts to make a "prediction of [the merger's] impact upon competition conditions in the future," which cannot be known beforehand. Even where an enforcement action seeks to break up an already consummated merger, the relevant analysis must predict what the world would look like had the merger never happened, something that cannot be known with certainty. See: Mark J. Niefer and Aaron D. Hoag, "Artificial Intelligence, Uncertainty and Merger Review," 220, citing United States v. Philadelphia National Bank, 374 U.S. 321, 362 (1963).

[113] For instance, the FTC was criticized by a federal district court judge in 2023 for its attempt to block the merger of cancer test producer GRAIL and medical test platform Illumina, as it failed to consider contractual assurances that the parties offered to make GRAIL's test available to other medical test platforms and thus mitigate anticompetitive concerns about the deal. Despite this, and under the mounting costs and pressure of litigation, the parties abandoned the deal. This likely resulted in the delayed rollout of life-saving cancer detection technology to patients as GRAIL lost out on Illumina's ability to expediently secure regulatory clearances and opportunity to use its platform. "[T]he FTC's placement of theoretical harms above actual competitive

In that light, consider the policing of data-driven anticompetitive conduct under current American antitrust law, which implements a flexible "rule of reason" framework. This framework weighs legitimate business justifications or procompetitive benefits of a business practice against its anticompetitive harms on a case-by-case basis to block practices that are net harmful.[114] Except for some presumptively anticompetitive practices like rigging bids and price-fixing that are deemed illegal, American antitrust law punishes unreasonable restraints of trade under Sherman Act Section 1 and exclusionary conduct or attempted monopolization under Sherman Act Section 2 through the "rule of reason."[115] This framework offers courts the flexibility to tailor rulings and remedies to the circumstances rather than imposing bright-line rules that could inadvertently stymie competitive or pro-innovation business practices. Similarly, proposed mergers are illegal under Clayton Act Section 7 when they may substantially lessen competition or tend to create a monopoly.[116] This occurs when mergers enhance a firm's incentive and ability to violate Sherman Act Section 1 or Section 2, either unilaterally (such as by restricting rivals' access to a key input such as data) or through restrictive agreements with other firms to achieve the same outcome.[117] Firms that expand market share or gain a monopoly through a superior product, business acumen, or luck won't be penalized under these laws.[118] Those that engage in practices that cannot be justified by any other reason besides excluding rivals from competing will attract punishment.[119] When a business practice or merger carries both procompetitive and anticompetitive implications, enforcers can seek binding consent decrees from the parties that limit the practice or that remedy anticompetitive concerns as a condition of settling the case or approving the merger.[120] This

---

and consumer harms achieved the opposite of the agency's supposed mandate. The litigation reduced competition, hurting a smaller firm that faces considerable regulatory hurdles, and harming consumers, potentially costing the lives of many patients who might otherwise have had timely access to [cancer detection] tests." See: Satya Marar, "Unintended Consequences: The Real Effects of Populist Antitrust Policies on Competition and Small Enterprises," (Mercatus Policy Brief, Mercatus Center at George Mason University, May 14, 2024), https://ssrn.com/abstract=4898015 citing Illumina, Inc. v. Federal Trade Commission, No. 23-60167, slip op. (5th Cir. December 15, 2023); Dave Michaels, "Court Sides with FTC Finding Illumina-Grail Deal Anticompetitive," *Wall Street Journal*, December 16, 2023; Alden Abbott, "When Bad Antitrust Costs Lives: The Illumina/GRAIL Tragedy," *Truth on the Market*, April 4, 2023; Natallie Rocha, "Illumina Will Divest GRAIL after Losing Battle with FTC. How Did the San Diego Biotech's $7B Deal Unravel?," *San Diego Union-Tribune*, December 17, 2023.

[114] Marar, "Artificial Intelligence and Antitrust Law: A Primer," https://ssrn.com/abstract=4745321, 8. The Supreme Court describes the "rule of reason" analysis framework thusly: "[T]he plaintiff has the initial burden to prove that the challenged restraint has a substantial anticompetitive effect that harms consumers in the relevant market . . . if the plaintiff carries its burden, then the burden shifts to the defendant to show a procompetitive rationale for the restraint . . . [i]f the defendant makes this showing, then the burden shifts back to the plaintiff to demonstrate that the procompetitive efficiencies could be reasonably achieved through less anticompetitive means." See Ohio v. Am. Express Co., 138 S. Ct. 2274, 201 L. Ed. 2d 678, 2284 (2018). An "anticompetitive effect" is harm to the competitive process, not simply a negative impact on a single market performance component (such as increased price). See: Gregory J. Werden, *The Foundations of Antitrust* (Durham, NC: Carolina Academic Press, 2020), 246–47 cited in Marar, "Artificial Intelligence and Antitrust Law: A Primer," 9.

[115] Alden F. Abbott, et al. "102 Antitrust Law Primer for Small Department Practitioners," 46-47.

[116] Marar, "Artificial Intelligence and Antitrust Law: A Primer," 8.

[117] See Stuart D. Gurrea and Bruce M. Owen, "Coordinated Interaction and Clayton Enforcement," *George Mason Law Review*, 12 (2003): 89.

[118] United States v. Grinnell Corp., 384 U.S. 563 (1966).

[119] The alleged monopolist must be shown to have harmed the competitive process itself through its actions, not simply to have harmed one or more of its competitors. See United States of America v. Microsoft Corporation, 253 F.3d 34 (D.C. Cir. 2001). (*'Microsoft'*) at 58.

[120] For instance, if the potential harms flowing from a merger between a large digital platform and a small software producer are that the platform will limit the availability of the smaller producer's software to their own platform and thus restrict its availability to other platforms and their users, then the parties can sign a binding decree whereby the merged firm agrees to make

process ensures a faster and more cost-effective resolution of complaints while giving businesses the freedom to experiment.

Antitrust claims under the rule of reason rest on legal arguments about how a merger or business practice may harm competition and consumers, known as theories of harm. Three theories of harm recognized by the US Supreme Court under the antitrust laws could be applied to a company's exclusionary conduct or attempted monopolization using data or foundation models trained on data. These are (1) raising rivals' costs, (2) refusing to deal with rivals, and (3) tying or bundling.

A dominant generative AI firm would raise the costs to compete incurred by its rivals and would thus be liable for monopolization if it acquires exclusive rights to a proprietary dataset or foundation model from a downstream supplier (or acquires the supplier itself) with the intent to restrict access for competitors, thus harming competition. The defendant would not be liable for exclusionary conduct or attempted monopolization if they could show procompetitive justifications for the practice,[121] or if the alleged key input could be substituted (such as by sourcing an alternate supplier) at low or no cost.[122] Legitimate justifications could include investing in the development and refinement of the model, in further data gathering, or investing in developing the platform through which it makes the model or dataset available.

Similarly, while firms can normally freely choose who they deal with and under what terms,[123] a firm can be held liable for "refusal to deal" if they sell a key input over which they hold a monopoly (such as a specific dataset that cannot be reasonably replicated) to another firm without a legitimate business justification.[124] This would be the case where the monopolist firm is shown to have sacrificed short-term profits for the sake of a long-term anticompetitive end.[125] Such cases are generally hard to win for plaintiffs relative to raising rivals' costs claims. However, refusal-to-deal claims have better prospects of success when the defendant firm has previously abandoned procompetitive cooperation with its rival or rivals. This behavior gives courts a baseline from which to judge the refusal-to-deal's anticompetitive effects and intent. Courts may also be reluctant to mandate that a firm must deal with its rivals because administering the duty to deal and setting its terms of dealing (such as price) can be a difficult exercise they may lack the expertise to undertake.[126] Thus, enforcers are likely to have greater success in prosecuting the same cases based on the raising rivals' costs theory.

---

the software available to other platforms post-merger. Similarly, in generative AI, a foundation model producer that acquires a company that produces open-source datasets for model training can consent to keeping the datasets available on an open-source basis as a condition for regulators to permit the acquisition. Such decrees entail costs for enforcers in monitoring to ensure compliance. However, they also promote competition and innovation by allowing procompetitive deals while mitigating any potential harms.

[121] Verizon Communications Inc. v. Law Offices of Curtis V. Trinko, LLP, 540 U.S. 398, 408 n.3 (2004). *('Trinko');* LePage's Inc. v. 3M, 324 F.3d 141, 150 (3d Cir. 2003); Aspen Highlands Skiing Corp. v. Aspen Skiing Co., 738 F.2d 1509 (10th Cir. 1984).

[122] Thomas Krattenmaker and Steven Salop. "Anticompetitive Exclusion: Raising Rivals' Costs to Achieve Power over Price," *Yale LJ* 96 (1986): 230.

[123] United States v. Trans-Mo. Freight Ass'n, 166 U.S. 290, 320 (1897).

[124] Verizon Communications Inc. v. Law Offices of Curtis V. Trinko, LLP, 540 U.S. 398, 408 n.3 (2004). *('Trinko');* LePage's Inc. v. 3M, 324 F.3d 141, 150 (3d Cir. 2003); Aspen Highlands Skiing Corp. v. Aspen Skiing Co., 738 F.2d 1509 (10th Cir. 1984).

[125] *Trinko* at 398.

[126] Ibid at 407, 408.

In some instances, if a firm were to condition the purchase of a product over which it has "market power" on the purchase of another one of its products, such as by only selling them together, that would constitute illegal exclusionary conduct.[127] Such arrangements are known as "tying" or "bundling" and generally have legitimate business justifications.[128] For instance, it makes sense to bundle multiple software applications on an operating system if they integrate and work more effectively together or if the bundled applications function as a single software 'ecosystem' with an integrated user experience under the operating system.[129] Similarly, firms that bundle foundation models or datasets with software applications, other datasets, or additional tools can be expected to argue that they are selling an integrated ecosystem of products that deliver enhanced functionality (or other benefits like data security) for users. They may argue that refinement of the model through additional feedback requires or is facilitated by bundling multiple applications built off the model. They may also argue that additional profits from the sales of the tied product help fund improvements in the tying product. For instance, the exorbitant additional costs of further data gathering and fine-tuning through machine learning might be funded through sales of additional software or web tools that integrate with the foundation model. In the case of an open-source foundation model, the provider may argue that requiring developers who build applications off the platform to license additional software before integrating their applications into the main platform may be necessary to be able to provide the model's source code for free.[130] Regardless, if a proprietary dataset or foundation model is sufficiently unique to grant the provider market power, then a sales arrangement that forces users or application developers to make additional purchases without any legitimate business purpose may be anticompetitive monopolization. Allowing courts to make such determinations without bright-line

---

[127] See United States v. Loew's, Inc., 371 U.S. 38 (1962). A tying or bundling arrangement may be deemed illegal if the defendant has "market power" over one of the products; if they condition its purchase on the purchase of another product; and if the arrangement affects a significant volume of interstate commerce. See: Jefferson Parish Hospital District No. 2 v. Hyde, 466 U.S. 2, 12-8 (1984). Even if these elements are met, some courts have required plaintiffs to show additional anticompetitive effects, while others will consider procompetitive justifications for the practice. See: Wells Real Estate, Inc. v. Greater Lowell Board of Realtors, 850 F.2d 803, 815 (1st Cir. 1988); Mozart Co. v. Mercedes-Benz of North America, Inc., 833 F.2d 1342, 1348–51 (9th Cir. 1987). Establishing that a defendant has "market power" over the tying product also entails showing that they can "raise prices or to require purchasers to accept burdensome terms that could not be exacted in a completely competitive market." See: United States Steel Corp. v. Fortner Enterprises, 429 U.S. 610 (1977). This test isn't satisfied by simply showing that a different price for the product can be charged to different customer segments. See: Michael E. Levine, "Price Discrimination Without Market Power," *Yale Journal on Regulation* 19, no. 1 (2002): 8.

[128] Courts have found that most tying arrangements are "fully consistent with a free, competitive market," and have recognized that most tie ins foster competition, even when the defendant has tying product power. See: Illinois Tool Works Inc. v. Independent Ink, Inc., 126 S. Ct. 1281, 1288 (2006); Phillip E. Areeda and Herbert Hovenkamp, *Antitrust Law: An Analysis of Antitrust Principles and Their Application*, 2nd ed. (Philadelphia, PA: Wolters Kluwer, 2004), section 518 at 220.

[129] "Since the utility of platform software (such as a computer or smartphone operating system) is providing a range of different applications and services in one place, it would be inappropriate to assume that the platform and each software component are separate products tied together, rather than a single offering that benefits consumers through convenience, a better user experience, and a reduction in time and resource expenditure relative to independently sourcing or distributing different services and applications. Integrating new functionalities into existing software is an in-demand and sought-after innovation for consumers, as consumers often buy into an entire 'ecosystem' of services and applications rather than individual products. Even when the two tied items are considered separate products, technologically and physically integrating them could improve the value of one or both individual products to users as well as to the producer of the complementary product." See: Satya Marar, "Artificial Intelligence and Antitrust Law: A Primer," 23, citing *Microsoft.*

[130] The same justification could be given for requiring integration into the main platform as a condition for using an open-source model's code for commercial purposes.

prohibitions on potentially procompetitive tying and bundling arrangements facilitates innovation and competition.

The three theories of harm recognized by US courts under existing antitrust law provide a strong legal farmwork to hold wrongdoers liable for anticompetitive abuses, such as abuses of control over data in generative AI markets. This framework functions without requiring preemptive interventions or new regulations. Importantly, it also allows businesses the flexibility to experiment with and justify procompetitive practices.

Conclusion

Current evidence and trends indicate that a thriving market of open-source and licensable proprietary foundation models and datasets will continue to foster competition in the generative AI space while mitigating concerns about anticompetitive behavior. Evidence from the current and foreseeable state of the generative AI market indicate that these concerns have not played out. Instead, trends and evidence indicate that tech giants and large digital platforms possess no inherent unassailable competitive advantage due to their unique access to data. Indeed, commercial arrangements between these giants and smaller firms are mutually beneficial, driving innovation and competition by combining their respective synergies and granting start-ups access to capital investment. Technological innovations, such as synthetic data, are also likely to further erode potential competitive barriers by reducing the costs of gathering and using data. These developments support the case against preemptive antitrust enforcement or bright-line competition rules that limit business conduct in the name of mitigating concerns about data-driven anticompetitive behavior. Such rules would hinder innovation and technological development in the AI space, threatening to cause significant economic loss in key innovation-driven sectors where the United States competes with China.

In contrast to potential new rules, including bright-line prohibitions, existing antitrust law offers ample means of prosecuting anticompetitive conduct, including attempts to monopolize or exclude rivals from competing. By applying a flexible and pragmatic 'rule of reason' framework, antitrust courts in the United States can judge business practices and proposed mergers on a case-by-case basis, weighing anticompetitive implications against procompetitive effects or legitimate business justifications for the deal or practice. As generative AI is a rapidly developing space impacting rapidly evolving markets in every economic sector, competition enforcers should continue to closely monitor and study these markets to identify changes or trends that could affect competitive conditions. Any regulatory or legal reforms to address future problems as they emerge should undergo a rigorous cost-benefit analysis prior to being undertaken.