

Comments of the International Center for Law & Economics

CMA Invitation to Comment on AI Partnerships

9 May 2024

Authored by:

Geoffrey A. Manne (President and Founder, International Center for Law & Economics)

Dirk Auer (Director of Competition Policy, International Center for Law & Economics)

Mario A. Zuniga (Senior Competition Policy Scholar, International Center for Law & Economics)

Executive Summary

We thank the Competition and Markets Authority (CMA) for this invitation to comment (ITC) on partnerships and other arrangements involving artificial intelligence (AI).¹ The International Center for Law & Economics (ICLE) is a nonprofit, nonpartisan global research and policy center founded with the goal of building the intellectual foundations for sensible, economically grounded policy. ICLE promotes the use of law & economics methodologies to inform public-policy debates and has longstanding expertise in the evaluation of competition law and policy. ICLE's interest is to ensure that competition law remains grounded in clear rules, established precedent, a record of evidence, and sound economic analysis.

In our comments, we express concern that policymakers' current concerns about competition in AI industries may be unwarranted. This is particularly true of the notion that incumbent digital platforms may use strategic partnerships with AI firms to insulate themselves from competition, including the three transactions that are central to the current ITC:

1. Amazon's partnership with Anthropic;
2. Microsoft's partnership with Mistral AI; and,
3. Microsoft's hiring of former Inflection AI employees (including, notably, founder Mustafa Suleyman) and related arrangements with the company.

Indeed, publicly available information suggests these transactions may not warrant merger-control investigation, let alone the heightened scrutiny that comes with potential Phase II proceedings. At the very least, given the AI industry's competitive landscape, there is little to suggest these transactions merit closer scrutiny than similar deals in other sectors.

Overenforcement in the field of generative AI paradoxically could engender the very harms that policymakers currently seek to avert. As we explain in greater detail below, preventing so-called "big tech" firms from competing in these markets (for example, by threatening competition intervention as soon as they build strategic relationships with AI startups) may thwart an important source of competition needed to keep today's leading generative-AI firms in check. In short, competition in AI markets is important², but trying naïvely to hold incumbent (in adjacent markets) tech firms back out of misguided fears they will come to dominate this space is likely to do more harm than good.

At a more granular level, there are important reasons to believe these agreements will have no negative impact on competition and they may, in fact, benefit consumers—*e.g.*, by enabling those startups to raise capital and deploy their services at an even larger scale. In other words, they do not bear any

¹ CMA Seeks Views on AI Partnerships and Other Arrangements, COMPETITION AND MARKETS AUTHORITY (24 Apr. 2024), <https://www.gov.uk/government/news/cma-seeks-views-on-ai-partnerships-and-other-arrangements>.

² AI, of course, is not a market (at least not a relevant antitrust market). Within the realm of what is being called "AI", companies can offer myriad products and services, and specific relevant markets would need to be defined before assessing harm to competition in specific cases.

of the *prima facie* traits of “killer acquisitions” or even of the acquisition of “nascent potential competitors”.³

Most importantly, these partnerships all involve the acquisition of minority stakes that do not entail any change of control over the target companies. Amazon, for instance, will not have “ownership control” of Anthropic. The precise amount of shares acquired has not been made public, but a reported investment of \$4 billion in a company valued at \$18.4 billion does not give Amazon a majority stake or sufficient voting rights to control the company or its competitive strategy.⁴ It has also been reported that the deal will not give Amazon any seats on the Anthropic board or special voting rights (such as the power to veto some decisions).⁵ There is thus little reason to believe Amazon has acquired indirect or *de facto* control over Anthropic.

Microsoft’s investment in Mistral AI is even smaller, in both absolute and relative terms. Microsoft is reportedly investing only \$16 million in a company valued at \$2.1 billion.⁶ This represents less than 1% of Mistral’s equity, making it all but impossible for Microsoft to exert any significant control or influence over Mistral AI’s competitive strategy. Likewise, there have been no reports of Microsoft acquiring seats on Mistral AI’s board or special voting rights. We can therefore be confident that the deal will not affect competition in AI markets.

Much of the same applies to Microsoft’s dealings with Inflection AI. Microsoft hired two of the company’s three founders (which currently does not fall under the scope of merger laws), and also paid \$620 million for nonexclusive rights to sell access to the Inflection AI model through its Azure Cloud.⁷ Admittedly, the latter could entail (depending on deal’s specifics) some limited control over Inflection AI’s competitive strategy, but there is currently no evidence to suggest this will be the case.

Finally, none of these deals entails any competitively significant behavioral commitments from the target companies. There are no reports of exclusivity agreements or other commitments that would

³ OECD, *Start-ups, Killer Acquisitions and Merger Control* (2020), available at <https://web.archive.oecd.org/2020-10-16/566931-start-ups-killer-acquisitions-and-merger-control-2020.pdf>.

⁴ Kate Rooney & Hayden Field, *Amazon Spends \$2.75 Billion on AI Startup Anthropic in Its Largest Venture Investment Yet*, CNBC (27 Mar. 2024), <https://www.cnbc.com/2024/03/27/amazon-spends-2point7b-on-startup-anthropic-in-largest-venture-investment.html>.

⁵ *Id.*

⁶ Tom Warren, *Microsoft Partners with Mistral in Second AI Deal Beyond OpenAI*, THE VERGE (26 Feb. 2024), <https://www.theverge.com/2024/2/26/24083510/microsoft-mistral-partnership-deal-azure-ai>.

⁷ Mark Sullivan, *Microsoft’s Inflection AI Grab Likely Cost More Than \$1 Billion, Says An Insider (Exclusive)*, FAST COMPANY (26 Mar. 2024), <https://www.fastcompany.com/91069182/microsoft-inflection-ai-exclusive>; see also, Mustafa Suleyman, *DeepMind and Inflection Co-Founder, Joins Microsoft to Lead Copilot*, MICROSOFT CORPORATE BLOGS (19 Mar. 2024), <https://blogs.microsoft.com/blog/2024/03/19/mustafa-suleyman-deepmind-and-inflection-co-founder-joins-microsoft-to-lead-copilot>; Krystal Hu & Harshita Mary Varghese, *Microsoft Pays Inflection \$ 650 Mln in Licensing Deal While Poaching Top Talent, Source Says*, REUTERS (21 Mar. 2024), <https://www.reuters.com/technology/microsoft-agreed-pay-inflection-650-mln-while-hiring-its-staff-information-2024-03-21>; *The New Inflection: An Important Change to How We’ll Work*, INFLECTION (Mar. 19, 2024), <https://inflection.ai/the-new-inflection>; Julie Bort, *Here’s How Microsoft Is Providing a ‘Good Outcome’ for Inflection AI VCs, as Reid Hoffman Promised*, TECH CRUNCH (21 Mar. 2024), <https://techcrunch.com/2024/03/21/microsoft-inflection-ai-investors-reid-hoffman-bill-gates>.

restrict third parties' access to these firms' underlying AI models. Again, this means the deals are extremely unlikely to negatively impact the competitive landscape in these markets.

At a more macro level, how the CMA deals with these proposed partnerships could have important ramifications for the UK economy. On the one hand, competition authorities (including the CMA) may be tempted to avoid the mistakes they arguably made during the formative years of what have become today's largest online platforms.⁸ The argument is that tougher enforcement may have reduced the high levels of concentration we see in these markets (the counterpoint is that these markets present features that naturally lead to relatively high levels of concentration and that this concentration benefits consumers in several ways⁹).

Unfortunately, this urge to curtail false negatives may come at the expense of judicial errors that hobble the UK economy. Discussing the EU's AI Act during a recent interview, French President Emmanuel Macron implicitly suggested the UK is in a unique position to attract AI (and other tech) investments away from the European Union. In his words:

We can decide to regulate much faster and much stronger than our major competitors. But we will regulate things that we will no longer produce or invent. This is never a good idea...

When I look at France, it is probably the first country in terms of artificial intelligence in continental Europe. We are neck and neck with the British. They will not have this regulation on foundational models. But above all, we are all very far behind the Chinese and the Americans.¹⁰

To capitalise on this opportunity, however, the UK must foster a fertile environment for startup activity. The CMA's approach to merger review in the AI industry is a small, but important, part of this picture. Looking at AI partnerships in an even-handed manner would signal a commitment to evidence-based policymaking that creates legal certainty for startups. For instance, sound merger-review principles would assure founders that corporate acquisition will remain a viable exit strategy in all but exceptional circumstances.

⁸ See Rana Foroohar, *The Great US-Europe Antitrust Divide*, FINANCIAL TIMES (5 Feb. 2024),

<https://www.ft.com/content/065a2f93-dc1e-410c-ba9d-73c930cedc14> (quoting FTC Chair Lina Khan "we are still reeling from the concentration that resulted from Web 2.0, and we don't want to repeat the mis-steps of the past with AI").

⁹ See, e.g., Geoffrey Manne & Dirk Auer, *Antitrust Dystopia and Antitrust Nostalgia: Alarmist Theories of Harm in Digital Markets and Their Origins*, 28 GEO. MASON L. REV. 1279, 1294 (2021). ("But while these increasing returns can cause markets to become more concentrated, they also imply that it is often more efficient to have a single firm serve the entire market. For instance, to a first approximation, network effects, which are one potential source of increasing returns, imply that it is more valuable—not just to the platform, but to the users themselves—for all users to be present on the same network or platform. In other words, fragmentation—de-concentration—may be more of a problem than monopoly in markets that exhibit network effects and increasing returns to scale. Given this, it is far from clear that antitrust authorities should try to prevent consolidation in markets that exhibit such characteristics, nor is it self-evident that these markets somehow produce less consumer surplus than markets that do not exhibit such increasing returns".)

¹⁰ Javier Espinoza & Leila Abboud, *EU's New AI Act Risks Hampering Innovation, Warns Emmanuel Macron*, FINANCIAL TIMES (11 Dec. 2023), <https://www.ft.com/content/9339d104-7b0c-42b8-9316-72226dd4e4c0>.

Of course, none of this is to say that established competition-law principles should play second fiddle to broader geopolitical ambitions. It does, however, suggest that the cost of false positives is particularly high in key industries like AI.

In short, how the CMA approaches these AI partnerships is of pivotal importance for both UK competition policy and the country's broader economic ambitions. The CMA should therefore look at these partnerships with an open mind, despite the important political and reputational pressure to be seen as “doing something” in this cutting-edge industry. Generative AI is already changing the ways that many firms do business and improving employee productivity in many industries.¹¹ The technology is also increasingly useful in the field of scientific research, where it has enabled new complex models that expand scientists' reach.¹² And while sensible enforcement is of vital importance to maintain competition and consumer welfare, it must be grounded in empirical evidence.

In the remainder of these comments, we will discuss the assumptions that underpin calls for heightened competition scrutiny in AI industries, and explain why they are unfounded. The big picture is that AI markets have grown rapidly, and new players are thriving. This would suggest that competition is intense. If incumbent firms could easily leverage their dominance into burgeoning generative-AI markets, we would not have seen the growth of such AI “unicorns” as OpenAI, Midjourney, and Anthropic, to name but a few. Furthermore, AI platforms developed by incumbent data collectors—such as Meta's Llama or Google's Bard, recently relaunched as Gemini—have struggled to gain traction. Of course, this is not to say that competition enforcers shouldn't care about generative AI markets, but rather that there is currently no apparent need for increased competition scrutiny in these markets.

The comments proceed as follows. Section I summarises recent calls for competition intervention in generative-AI markets. Section II argues that many of these calls are underpinned by fears of data-related incumbency advantages (often referred to as “data-network effects”), including in the context of mergers. Section III explains why these effects are unlikely to play a meaningful role in generative-AI markets. Section IV concludes by offering five key takeaways to help policymakers better weigh the tradeoffs inherent to competition intervention (including merger control) in generative-AI markets.

¹¹ See, e.g., Michael Chui, et al., *The Economic Potential of Generative AI: The Next Productivity Frontier*, MCKINSEY (14 Jun. 2023), <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier>.

¹² See, e.g., Zhuoran Qiao, Weili Nie, Arash Vahdat, Thomas F. Miller III, & Animashree Anandkumar, *State-Specific Protein-Ligand Complex Structure Prediction with a Multiscale Deep Generative Model*, 6 NATURE MACHINE INTELLIGENCE, 195-208 (2024); see also Jaemin Seo, Sang Kyeun Kim, Azarakhsh Jalalvand, Rory Conlin, Andrew Rothstein, Joseph Abbate, Keith Erickson, Josiah Wai, Ricardo Shousha, & Egemen Kolemen, *Avoiding Fusion Plasma Tearing Instability with Deep Reinforcement Learning*, 626 NATURE, 746-751 (2024).

I. Calls for Intervention in AI Markets

It was once (and frequently) said that Google’s “data monopoly” was unassailable: “If ‘big data’ is the oil of the information economy, Google has Standard Oil-like monopoly dominance—and uses that control to maintain its dominant position”.¹³ Similar claims of data dominance have been attached to nearly all large online platforms, including Facebook (Meta), Amazon, and Uber.¹⁴

While some of these claims continue even today (for example, “big data” is a key component of the U.S. Justice Department’s (DOJ) Google Search and adtech antitrust suits),¹⁵ a shiny new data target has emerged in the form of generative artificial intelligence (AI). The launch of ChatGPT in November 2022, as well as the advent of AI image-generation services like Midjourney and Dall-E, have dramatically expanded the public’s conception of what is—and what might be—possible to achieve with generative-AI technologies built on massive datasets.

While these services remain both in the early stages of mainstream adoption and in the throes of rapid, unpredictable technological evolution, they nevertheless already appear to be on the radar of competition policymakers around the world. Several antitrust enforcers appear to believe that, by acting now, they can avoid the “mistakes” that were purportedly made during the formative years of Web 2.0.¹⁶ These mistakes, critics assert, include failing to appreciate the centrality of data in online markets, as well as letting mergers go unchecked and allowing early movers to entrench their market positions.¹⁷ As Lina Khan, chair of the U.S. Federal Trade Commission (FTC), put it: “we are still

¹³ Nathan Newman, *Taking on Google’s Monopoly Means Regulating Its Control of User Data*, HUFFINGTON POST (24 Sep. 2013), http://www.huffingtonpost.com/nathan-newman/taking-on-googlesmonopol_b_3980799.html.

¹⁴ See, e.g., Lina Khan & K. Sabeel Rahman, *Restoring Competition in the U.S. Economy*, in UNTAMED: HOW TO CHECK CORPORATE, FINANCIAL, AND MONOPOLY POWER (Nell Abernathy, Mike Konczal, & Kathryn Milani, eds., 2016), at 23 (“From Amazon to Google to Uber, there is a new form of economic power on display, distinct from conventional monopolies and oligopolies..., leverag[ing] data, algorithms, and internet-based technologies... in ways that could operate invisibly and anticompetitively”.); Mark Weinstein, *I Changed My Mind—Facebook Is a Monopoly*, WALL ST. J. (1 Oct. 2021), <https://www.wsj.com/articles/facebook-is-monopoly-metaverse-users-advertising-platforms-competition-mewe-big-tech-11633104247> (“[T]he glue that holds it all together is Facebook’s monopoly over data.... Facebook’s data troves give it unrivaled knowledge about people, governments—and its competitors”).

¹⁵ See, generally, Abigail Slater, *Why “Big Data” Is a Big Deal*, THE REG. REV. (6 Nov. 2023), <https://www.theregreview.org/2023/11/06/slater-why-big-data-is-a-big-deal>; Amended Complaint at ¶36, *United States v. Google*, 1:20-cv-03010- (D.D.C. 2020); Complaint at ¶37, *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> (“Google intentionally exploited its massive trove of user data to further entrench its monopoly across the digital advertising industry”).

¹⁶ See, e.g., Press Release, *Commission Launches Calls for Contributions on Competition in Virtual Worlds and Generative AI*, EUROPEAN COMMISSION (9 Jan. 2024), https://ec.europa.eu/commission/presscorner/detail/en/IP_24_85; Krysten Crawford, *FTC’s Lina Khan Warns Big Tech over AI*, SIEPR (3 Nov. 2020), <https://siepr.stanford.edu/news/ftcs-lina-khan-warns-big-tech-over-ai> (“Federal Trade Commission Chair Lina Khan delivered a sharp warning to the technology industry in a speech at Stanford on Thursday: *Antitrust enforcers are watching what you do in the race to profit from artificial intelligence*”). (emphasis added).

¹⁷ See, e.g., John M. Newman, *Antitrust in Digital Markets*, 72 VAND. L. REV. 1497, 1501 (2019) (“[T]he status quo has frequently failed in this vital area, and it continues to do so with alarming regularity. The laissez-faire approach advocated for by scholars and adopted by courts and enforcers has allowed potentially massive harms to go unchecked”.); Bertin Martins,

reeling from the concentration that resulted from Web 2.0, and we don't want to repeat the missteps of the past with AI".¹⁸

This response from the competition-policy world is deeply troubling. Rather than engage in critical self-assessment and adopt an appropriately restrained stance, the enforcement community appears to be champing at the bit. Rather than assessing their prior assumptions based on the current technological moment, enforcers' top priority appears to be figuring out how to rapidly and almost reflexively deploy existing competition tools to address the presumed competitive failures presented by generative AI.¹⁹

It is increasingly common for competition enforcers to argue that so-called "data-network effects" serve not only to entrench incumbents in those markets where the data is collected, but also to confer similar, self-reinforcing benefits in adjacent markets. Several enforcers have, for example, prevented large online platforms from acquiring smaller firms in adjacent markets, citing the risk that they could use their vast access to data to extend their dominance into these new markets.²⁰

They have also launched consultations to ascertain the role that data plays in AI competition. For instance, in an ongoing consultation, the European Commission asks: "What is the role of data and what are its relevant characteristics for the provision of generative AI systems and/or components, including AI models?"²¹ Unsurprisingly, the FTC has likewise been bullish about the risks posed by incumbents' access to data. In comments submitted to the U.S. Copyright Office, for example, the FTC argued that:

The rapid development and deployment of AI also poses potential risks to competition. The rising importance of AI to the economy may further lock in the market dominance of large incumbent technology firms. These powerful, vertically integrated incumbents control many of the inputs necessary for the effective development and deployment of AI tools, including cloud-based or local computing power *and access to large stores of training data*. These dominant technology companies may have the incentive to use their

Are New EU Data Market Regulations Coherent and Efficient?, Bruegel Working Paper 21/23 (2023), <https://www.bruegel.org/working-paper/are-new-eu-data-market-regulations-coherent-and-efficient> ("Technical restrictions on access to and re-use of data may result in failures in data markets and data-driven services markets".); Valéria Faure-Muntian, *Competitive Dysfunction: Why Competition Law Is Failing in a Digital World*, THE FORUM NETWORK (24 Feb. 2021), <https://www.oecd-forum.org/posts/competitive-dysfunction-why-competition-law-is-failing-in-a-digital-world>.

¹⁸ See Foroohar, *supra* note 8.

¹⁹ See, e.g., Press Release, European Commission, *supra* note 16.

²⁰ See *infra*, Section II. Commentators have also made similar claims; see, e.g., Ganesh Sitaram & Tejas N. Narechania, *It's Time for the Government to Regulate AI. Here's How*, POLITICO (15 Jan. 2024) ("All that cloud computing power is used to train foundation models by having them "learn" from incomprehensibly huge quantities of data. Unsurprisingly, the entities that own these massive computing resources are also the companies that dominate model development. Google has Bard, Meta has LLaMa. Amazon recently invested \$4 billion into one of OpenAI's leading competitors, Anthropic. And Microsoft has a 49 percent ownership stake in OpenAI – giving it extraordinary influence, as the recent board struggles over Sam Altman's role as CEO showed").

²¹ Press Release, European Commission, *supra* note 16.

control over these inputs to unlawfully entrench their market positions in AI and related markets, including digital content markets.²²

Certainly, it stands to reason that the largest online platforms—including Alphabet, Meta, Apple, and Amazon—should have a meaningful advantage in the burgeoning markets for generative-AI services. After all, it is widely recognised that data is an essential input for generative AI.²³ This competitive advantage should be all the more significant, given that these firms have been at the forefront of AI technology for more than a decade. Over this period, Google’s DeepMind and AlphaGo and Meta’s NLLB-200 have routinely made headlines.²⁴ Apple and Amazon also have vast experience with AI assistants, and all of these firms use AI technology throughout their platforms.²⁵

Contrary to what one might expect, however, the tech giants have, to date, been largely unable to leverage their vast data troves of data to outcompete startups like OpenAI and Midjourney. At the time of writing, OpenAI’s ChatGPT appears to be, by far, the most successful chatbot,²⁶ despite the large tech platforms’ apparent access to far more (and more up-to-date) data.

In these comments, we suggest that there are important lessons to glean from these developments, if only enforcers would stop to reflect. The meteoric rise of consumer-facing AI services should offer competition enforcers and policymakers an opportunity for *introspection*. As we explain, the rapid emergence of generative-AI technology may *undercut* many core assumptions of today’s competition-policy debates, which have focused largely on the rueful after-effects of the purported failure of 20th-century antitrust to address the allegedly manifest harms of 21st-century technology. These include the notions that data advantages constitute barriers to entry and can be leveraged to project

²² Comment of U.S. Federal Trade Commission to the U.S. Copyright Office, Artificial Intelligence and Copyright, Docket No. 2023-6 (30 Oct. 2023), at 4, <https://www.ftc.gov/legal-library/browse/advocacy-filings/comment-federal-trade-commission-artificial-intelligence-copyright> (emphasis added).

²³ See, e.g., Joe Caserta, Holger Harreis, Kayvaun Rowshankish, Nikhil Srinidhi, & Asin Tavakoli, *The Data Dividend: Fueling Generative AI*, MCKINSEY DIGITAL (15 Sep. 2023), <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-data-dividend-fueling-generative-ai> (“Your data and its underlying foundations are the determining factors to what’s possible with generative AI”).

²⁴ See, e.g., Tim Keary, *Google DeepMind’s Achievements and Breakthroughs in AI Research*, TECHOPEDIA (11 Aug. 2023), <https://www.techopedia.com/google-deepminds-achievements-and-breakthroughs-in-ai-research>; See, e.g., Will Douglas Heaven, *Google DeepMind Used a Large Language Model to Solve an Unsolved Math Problem*, MIT TECHNOLOGY REVIEW (14 Dec. 2023), <https://www.technologyreview.com/2023/12/14/1085318/google-deepmind-large-language-model-solve-unsolvable-math-problem-cap-set>; see also, *A Decade of Advancing the State-of-the-Art in AI Through Open Research*, META (30 Nov. 2023), <https://about.fb.com/news/2023/11/decade-of-advancing-ai-through-open-research>; see also, *200 Languages Within a Single AI Model: A Breakthrough in High-Quality Machine Translation*, META, <https://ai.meta.com/blog/nllb-200-high-quality-machine-translation> (last visited 18 Jan. 2023).

²⁵ See, e.g., Jennifer Allen, *10 Years of Siri: The History of Apple’s Voice Assistant*, TECH RADAR (4 Oct. 2021), <https://www.techradar.com/news/siri-10-year-anniversary>; see also Evan Selleck, *How Apple Is Already Using Machine Learning and AI in iOS*, APPLE INSIDER (20 Nov. 2023), <https://appleinsider.com/articles/23/09/02/how-apple-is-already-using-machine-learning-and-ai-in-ios>; see also, Kathleen Walch, *The Twenty Year History Of AI At Amazon*, FORBES (19 July 2019), <https://www.forbes.com/sites/cognitiveworld/2019/07/19/the-twenty-year-history-of-ai-at-amazon>.

²⁶ See *infra* Section III.

dominance into adjacent markets; that scale itself is a market failure to be addressed by enforcers; and that the use of consumer data is inherently harmful to those consumers.

II. Data-Network Effects Theory and Enforcement

Proponents of more extensive intervention by competition enforcers into digital markets often cite data-network effects as a source of competitive advantage and barrier to entry (though terms like “economies of scale and scope” may offer more precision).²⁷ The crux of the argument is that “the collection and use of data creates a feedback loop of more data, which ultimately insulates incumbent platforms from entrants who, but for their data disadvantage, might offer a better product”.²⁸ This self-reinforcing cycle purportedly leads to market domination by a single firm. Thus, it is argued, e.g., that Google’s “ever-expanding control of user personal data, and that data’s critical value to online advertisers, creates an *insurmountable barrier to entry for new competition*”.²⁹

But it is important to note the conceptual problems these claims face. Because data can be used to improve products’ quality and/or to subsidise their use, treating the possession 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. Of course, it would be a curious approach to antitrust if competition were treated as a problem, as it would imply that firms should under-compete—*i.e.*, should forego consumer-welfare enhancements—in order to inculcate a greater number of firms in a given market, simply for its own sake.³⁰

²⁷ See, e.g., Cédric Argenton & Jens Prüfer, *Search Engine Competition with Network Externalities*, 8 J. COMP. L. & ECON. 73, 74 (2012).

²⁸ John M. Yun, *The Role of Big Data in Antitrust*, in THE GLOBAL ANTITRUST INSTITUTE REPORT ON THE DIGITAL ECONOMY (Joshua D. Wright & Douglas H. Ginsburg, eds., 11 Nov. 2020) at 233, https://gaidigitalreport.com/2020/08/25/big-data-and-barriers-to-entry/#_ftnref50; see also, e.g., Robert Wayne Gregory, Ola Henfridsson, Evgeny Kaganer, & Harris Kyriakou, *The Role of Artificial Intelligence and Data Network Effects for Creating User Value*, 46 ACAD. OF MGMT. REV. 534 (2020), final pre-print version at 4, <http://wrap.warwick.ac.uk/134220> (“A platform exhibits data network effects if, the more that the platform learns from the data it collects on users, the more valuable the platform becomes to each user”.); see also, Karl Schmedders, José Parra-Moyano, & Michael Wade, *Why Data Aggregation Laws Could be the Answer to Big Tech Dominance*, SILICON REPUBLIC (6 Feb. 2024), <https://www.siliconrepublic.com/enterprise/data-ai-aggregation-laws-regulation-big-tech-dominance-competition-antitrust-imd>.

²⁹ Nathan Newman, *Search, Antitrust, and the Economics of the Control of User Data*, 31 YALE J. REG. 401, 409 (2014) (emphasis added); see also *id.* at 420 & 423 (“While there are a number of network effects that come into play with Google, [“its intimate knowledge of its users contained in its vast databases of user personal data”] is likely the most important one in terms of entrenching the company’s monopoly in search advertising.... Google’s overwhelming control of user data... might make its dominance nearly unchallengeable”).

³⁰ See also Yun, *supra* note 28 at 229 (“[I]nvestments in big data can create competitive distance between a firm and its rivals, including potential entrants, but this distance is the result of a competitive desire to improve one’s product”).

Meanwhile, actual economic studies of data-network effects have been few and far between, with scant empirical evidence to support the theory.³¹ Andrei Hagiu and Julian Wright’s theoretical paper offers perhaps the most comprehensive treatment of the topic to date.³² The authors ultimately conclude that data-network effects can be of differing magnitudes and have varying effects on firms’ incumbency advantage.³³ They cite Grammarly (an AI writing-assistance tool) as a *potential* example: “As users make corrections to the suggestions offered by Grammarly, its language experts and artificial intelligence can use this feedback to continue to improve its future recommendations for all users”.³⁴

This is echoed by other economists who contend that “[t]he algorithmic analysis of user data and information might increase incumbency advantages, creating lock-in effects among users and making them more reluctant to join an entrant platform”.³⁵ Crucially, some scholars take this logic a step further, arguing that platforms may use data from their “origin markets” in order to enter and dominate adjacent ones:

First, as we already mentioned, data 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. Second, data collected in the origin market can 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.³⁶

This possibility is also implicit in Hagiu and Wright’s paper.³⁷ Indeed, the authors’ theoretical model rests on an important distinction between within-user data advantages (that is, having access to more data about a given user) and across-user data advantages (information gleaned from having access to a wider user base). In both cases, there is an implicit assumption that platforms may use data from one service to gain an advantage in another market (because what matters is information about aggregate or individual user preferences, regardless of its origin).

³¹ For a review of the literature on *increasing returns to scale in data* (this topic is broader than data-network effects) see Geoffrey Manne & Dirk Auer, *Antitrust Dystopia and Antitrust Nostalgia: Alarmist Theories of Harm in Digital Markets and Their Origins*, 28 GEO MASON L. REV. 1281, 1344 (2021).

³² Andrei Hagiu & Julian Wright, *Data-Enabled Learning, Network Effects, and Competitive Advantage*, 54 RAND J. ECON. 638 (2023).

³³ *Id.* at 639. The authors conclude that “Data-enabled learning would seem to give incumbent firms a competitive advantage. But how strong is this advantage and how does it differ from that obtained from more traditional mechanisms...”.

³⁴ *Id.*

³⁵ Bruno Jullien & Wilfried Sand-Zantman, *The Economics of Platforms: A Theory Guide for Competition Policy*, 54 INFO. ECON. & POL’Y 10080, 101031 (2021).

³⁶ Daniele Condorelli & Jorge Padilla, *Harnessing Platform Envelopment in the Digital World*, 16 J. COMP. L. & POL’Y 143, 167 (2020).

³⁷ See Hagiu & Wright, *supra* note 32.

Our review of the economic evidence suggests that several scholars have, with varying degrees of certainty, raised the possibility that incumbents may leverage data advantages to stifle competitors in their primary market or in adjacent ones (be it via merger or organic growth). As we explain below, however, there is ultimately little evidence to support such claims. Policymakers have, however, been keenly receptive to these limited theoretical findings, basing multiple decisions on these theories, often with little consideration given to the caveats that accompany them.³⁸

Indeed, it is remarkable that, in its section on “[t]he data advantage for incumbents”, the “Furman Report” created for the UK government cited only two empirical economic studies, and they offer *directly contradictory conclusions* with respect to the question of the strength of data advantages.³⁹ Nevertheless, the Furman Report concludes that data “may confer a form of unmatched advantage on the incumbent business, making successful rivalry less likely”,⁴⁰ and adopts without reservation “convincing” evidence from non-economists that have no apparent empirical basis.⁴¹

In the *Google/Fitbit* merger proceedings, the European Commission found that the combination of data from Google services with that of Fitbit devices would reduce competition in advertising markets:

Giving [sic] the large amount of data already used for advertising purposes that Google holds, the increase in Google’s data collection capabilities, which goes beyond the mere number of active users for which Fitbit has been collecting data so far, the Transaction is likely to have a negative impact on the development of an unfettered competition in the markets for online advertising.⁴²

³⁸ For a summary of these limitations, see generally Catherine Tucker, *Network Effects and Market Power: What Have We Learned in the Last Decade?*, ANTITRUST (2018) at 72, available at <https://sites.bu.edu/tpri/files/2018/07/tucker-network-effects-antitrust2018.pdf>; see also Manne & Auer, *supra* note 31, at 1330.

³⁹ See Jason Furman, Diane Coyle, Amelia Fletcher, Derek McAuley, & Philip Marsden (Dig. Competition Expert Panel), *Unlocking Digital Competition* (2019) at 32-35 (“Furman Report”), available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/785547/unlocking_digital_competition_furman_review_web.pdf.

⁴⁰ *Id.* at 34.

⁴¹ *Id.* at 35. To its credit, it should be noted, the Furman Report does counsel caution before mandating access to data as a remedy to promote competition. See *id.* at 75. That said, the Furman Report maintains that such a remedy should certainly be on the table because “the evidence suggests that large data holdings are at the heart of the potential for some platform markets to be dominated by single players and for that dominance to be entrenched in a way that lessens the potential for competition for the market”. *Id.* In fact, the evidence does not show this.

⁴² Case COMP/M.9660 – Google/Fitbit, Commission Decision (17 Dec. 2020) (Summary at O.J. (C 194) 7), available at https://ec.europa.eu/competition/mergers/cases1/202120/m9660_3314_3.pdf at 455.

As a result, the Commission cleared the merger on the condition that Google refrain from using data from Fitbit devices for its advertising platform.⁴³ The Commission also appears likely to focus on similar issues in its ongoing investigation of Microsoft’s investment into OpenAI.⁴⁴

Along similar lines, in its complaint to enjoin Meta’s purchase of Within Unlimited—makers of the virtual-reality (VR) fitness app Supernatural—the FTC relied on, among other things, the fact that Meta could leverage its data about VR-user behavior to inform its decisions and potentially outcompete rival VR-fitness apps: “Meta’s control over the Quest platform also gives it unique access to VR user data, which it uses to inform strategic decisions”.⁴⁵

The DOJ’s twin cases against Google also implicate data leveraging and data barriers to entry. The agency’s adtech complaint charges that “Google intentionally exploited its massive trove of user data to further entrench its monopoly across the digital advertising industry”.⁴⁶ Similarly, in its Google Search complaint, the agency argues that:

Google’s anticompetitive practices are especially pernicious because they deny rivals scale to compete effectively. 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.⁴⁷

Finally, updated merger guidelines published in recent years by several competition enforcers cite the acquisition of data as a potential source of competition concerns. For instance, the FTC and DOJ’s 2023 guidelines state that “acquiring data that helps facilitate matching, sorting, or prediction services may enable the platform to weaken rival platforms by denying them that data”.⁴⁸ Likewise, the CMA itself warns against incumbents acquiring firms in order to obtain their data and foreclose other rivals:

Incentive to foreclose rivals...

⁴³ *Id.* at 896.

⁴⁴ See Natasha Lomas, *EU Checking if Microsoft’s OpenAI Investment Falls Under Merger Rules*, TECHCRUNCH (9 Jan. 2024), <https://techcrunch.com/2024/01/09/openai-microsoft-eu-merger-rules>.

⁴⁵ Amended Complaint at 11, *Meta/Zuckerberg/Within*, FED. TRADE COMM’N. (2022) (No. 605837), available at https://www.ftc.gov/system/files/ftc_gov/pdf/D09411%20-%20AMENDED%20COMPLAINT%20FILED%20BY%20COUNSEL%20SUPPORTING%20THE%20COMPLAINT%20-%20PUBLIC%20%281%29_0.pdf.

⁴⁶ Amended Complaint (D.D.C), *supra* note 15 at ¶137.

⁴⁷ Amended Complaint (E.D. Va), *supra* note 15 at ¶18.

⁴⁸ *Merger Guidelines*, US DEP’T OF JUSTICE & FED. TRADE COMM’N (2023) at 25, available at https://www.ftc.gov/system/files/ftc_gov/pdf/2023_merger_guidelines_final_12.18.2023.pdf.

7.19(e) Particularly in complex and dynamic markets, firms may not focus on short term margins but may pursue other objectives to maximise their long-run profitability, which the CMA may consider. This may include... obtaining access to customer data...⁴⁹

In short, competition authorities around the globe have been taking an increasingly aggressive stance on data-network effects. Among the ways this has manifested is in enforcement decisions based on fears that data collected by one platform might confer a decisive competitive advantage in adjacent markets. Unfortunately, these concerns rest on little to no empirical evidence, either in the economic literature or the underlying case records.

III. Data-Incumbency Advantages in Generative-AI Markets

Given the assertions detailed in the previous section, it would be reasonable to assume that firms such as Google, Meta, and Amazon should be in pole position to dominate the burgeoning market for generative AI. After all, these firms have not only been at the forefront of the field for the better part of a decade, but they also have access to vast troves of data, the likes of which their rivals could only dream when they launched their own services. Thus, the authors of the Furman Report caution that “to the degree that the next technological revolution centres around artificial intelligence and machine learning, then the companies most able to take advantage of it may well be the existing large companies because of the importance of data for the successful use of these tools”.⁵⁰

To date, however, this is not how things have unfolded—although it bears noting these markets remain in flux and the competitive landscape is susceptible to change. The first significantly successful generative-AI service was arguably not from either Meta—which had been working on chatbots for years and had access to, arguably, the world’s largest database of actual chats—or Google. Instead, the breakthrough came from a previously unknown firm called OpenAI.

OpenAI’s ChatGPT service currently holds an estimated 60% of the market (though reliable numbers are somewhat elusive).⁵¹ It broke the record for the fastest online service to reach 100 million users (in only a couple of months), more than four times faster than previous record holder TikTok.⁵² Based on Google Trends data, ChatGPT is nine times more popular worldwide than Google’s

⁴⁹ *Merger Assessment Guidelines*, COMPETITION AND MKTS. AUTH (2021) at ¶7.19(e), available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1051823/MAGs_for_publication_2021_-_pdf.

⁵⁰ Furman Report, *supra* note 39, at ¶4.

⁵¹ See, e.g., Chris Westfall, *New Research Shows ChatGPT Reigns Supreme in AI Tool Sector*, FORBES (16 Nov. 2023), <https://www.forbes.com/sites/chriswestfall/2023/11/16/new-research-shows-chatgpt-reigns-supreme-in-ai-tool-sector/?sh=7de5de250e9c>.

⁵² See Krystal Hu, *ChatGPT Sets Record for Fastest-Growing User Base*, REUTERS (2 Feb. 2023), <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01>; Google: *The AI Race Is On*, APP ECONOMY INSIGHTS (7 Feb. 2023), <https://www.appconomyinsights.com/p/google-the-ai-race-is-on>.

own Bard service, and 14 times more popular in the United States.⁵³ In April 2023, ChatGPT reportedly registered 206.7 million unique visitors, compared to 19.5 million for Google’s Bard.⁵⁴ In short, at the time we are writing, ChatGPT appears to be the most popular chatbot. The entry of large players such as Google Bard or Meta AI appear to have had little effect thus far on its market position.⁵⁵

The picture is similar in the field of AI-image generation. As of August 2023, Midjourney, Dall-E, and Stable Diffusion appear to be the three market leaders in terms of user visits.⁵⁶ This is despite competition from the likes of Google and Meta, who arguably have access to unparalleled image and video databases by virtue of their primary platform activities.⁵⁷

This raises several crucial questions: how have these AI upstarts managed to be so successful, and is their success just a flash in the pan before Web 2.0 giants catch up and overthrow them? While we cannot answer either of these questions dispositively, we offer what we believe to be some relevant observations concerning the role and value of data in digital markets.

A first important observation is that empirical studies suggest that data exhibits diminishing marginal returns. In other words, past a certain point, acquiring more data does not confer a meaningful edge to the acquiring firm. As Catherine Tucker put it following a review of the literature: “Empirically there is little evidence of economies of scale and scope in digital data in the instances where one would expect to find them”.⁵⁸

Likewise, following a survey of the empirical literature on this topic, Geoffrey Manne and Dirk Auer conclude that:

Available evidence suggests that claims of “extreme” returns to scale in the tech sector are greatly overblown. Not only are the largest expenditures of digital platforms unlikely

⁵³ See Google Trends, <https://trends.google.com/trends/explore?date=today%205-y&q=%2Fg%2F11khcfz0y2,%2Fg%2F11ts49p01g&hl=en> (last visited 12 Jan. 2024) and <https://trends.google.com/trends/explore?date=today%205-y&geo=US&q=%2Fg%2F11khcfz0y2,%2Fg%2F11ts49p01g&hl=en> (last visited 12 Jan. 2024).

⁵⁴ See David F. Carr, *As ChatGPT Growth Flattened in May, Google Bard Rose 187%*, SIMILARWEB BLOG (5 Jun. 2023), <https://www.similarweb.com/blog/insights/ai-news/chatgpt-bard>.

⁵⁵ See Press Release, *Introducing New AI Experiences Across Our Family of Apps and Devices*, META (27 Sep. 2023), <https://about.fb.com/news/2023/09/introducing-ai-powered-assistants-characters-and-creative-tools>; Sundar Pichai, *An Important Next Step on Our AI Journey*, GOOGLE KEYWORD BLOG (Feb. 6, 2023), <https://blog.google/technology/ai/bard-google-ai-search-updates>.

⁵⁶ See Ion Prodan, *14 Million Users: Midjourney’s Statistical Success*, YON (Aug. 19, 2023), <https://yon.fun/midjourney-statistics>; see also Andrew Wilson, *Midjourney Statistics: Users, Polls, & Growth [Oct 2023]*, APPROACHABLEAI (13 Oct. 2023), <https://approachableai.com/midjourney-statistics>.

⁵⁷ See Hema Budaraju, *New Ways to Get Inspired with Generative AI in Search*, GOOGLE KEYWORD BLOG (12 Oct. 2023), <https://blog.google/products/search/google-search-generative-ai-october-update>; *Imagine with Meta AI*, META (last visited Jan. 12, 2024), <https://imagine.meta.com>.

⁵⁸ Catherine Tucker, *Digital Data, Platforms and the Usual [Antitrust] Suspects: Network Effects, Switching Costs, Essential Facility*, 54 REV. INDUS. ORG. 683, 686 (2019).

to become proportionally less important as output increases, but empirical research strongly suggests that even data does not give rise to increasing returns to scale, despite routinely being cited as the source of this effect.⁵⁹

In other words, being the firm with the most data appears to be far less important than having *enough* data. Moreover, this lower bar may be accessible to far more firms than one might initially think possible. Furthermore, obtaining sufficient data could become easier still—that is, the volume of required data could become even smaller—with technological progress. For instance, synthetic data may provide an adequate substitute to real-world data,⁶⁰ or may even outperform real-world data.⁶¹ As Thibault Schrepel and Alex Pentland surmise:

[A]dvances in computer science and analytics are making the amount of data less relevant every day. In recent months, important technological advances have allowed companies with small data sets to compete with larger ones.⁶²

Indeed, past a certain threshold, acquiring more data might not meaningfully improve a service, where other improvements (such as better training methods or data curation) could have a large impact. In fact, there is some evidence that excessive data impedes a service’s ability to generate results appropriate for a given query: “[S]uperior model performance can often be achieved with smaller, high-quality datasets than massive, uncurated ones. Data curation ensures that training datasets are devoid of noise, irrelevant instances, and duplications, thus maximizing the efficiency of every training iteration”.⁶³

Consider, for instance, a user who wants to generate an image of a basketball. Using a model trained on an indiscriminate range and number of public photos in which a basketball appears surrounded by copious other image data, the user may end up with an inordinately noisy result. By contrast, a

⁵⁹ Manne & Auer, *supra* note 31, at 1345.

⁶⁰ See, e.g., Stefanie Koperniak, *Artificial Data Give the Same Results as Real Data—Without Compromising Privacy*, MIT NEWS (3 Mar. 2017), <https://news.mit.edu/2017/artificial-data-give-same-results-as-real-data-0303> (“[Authors] describe a machine learning system that automatically creates synthetic data—with the goal of enabling data science efforts that, due to a lack of access to real data, may have otherwise not left the ground. While the use of authentic data can cause significant privacy concerns, this synthetic data is completely different from that produced by real users—but can still be used to develop and test data science algorithms and models”).

⁶¹ See, e.g., Rachel Gordon, *Synthetic Imagery Sets New Bar in AI Training Efficiency*, MIT NEWS (20 Nov. 2023), <https://news.mit.edu/2023/synthetic-imagery-sets-new-bar-ai-training-efficiency-1120> (“By using synthetic images to train machine learning models, a team of scientists recently surpassed results obtained from traditional ‘real-image’ training methods.”).

⁶² Thibault Schrepel & Alex ‘Sandy’ Pentland, *Competition Between AI Foundation Models: Dynamics and Policy Recommendations*, MIT CONNECTION SCIENCE WORKING PAPER (Jun. 2023), at 8.

⁶³ Igor Susmelj, *Optimizing Generative AI: The Role of Data Curation*, LIGHTLY (last visited 15 Jan. 2024), <https://www.lightly.ai/post/optimizing-generative-ai-the-role-of-data-curation>.

model trained with a better method on fewer, more carefully selected images could readily yield far superior results.⁶⁴ In one important example:

[t]he model's performance is particularly remarkable, given its small size. "This is not a large language model trained on the whole Internet; this is a relatively small transformer trained for these tasks," says Armando Solar-Lezama, a computer scientist at the Massachusetts Institute of Technology, who was not involved in the new study.... The finding implies that instead of just shoving ever more training data into machine-learning models, a complementary strategy might be to offer AI algorithms the equivalent of a focused linguistics or algebra class.⁶⁵

Platforms' current efforts are thus focused on improving the mathematical and logical reasoning of large language models (LLMs), rather than maximizing training datasets.⁶⁶

Two points stand out. The first is that firms like OpenAI rely largely on *publicly available* datasets—such as GSM8K—to train their LLMs.⁶⁷ Second, the real challenge to create cutting-edge AI is not so much in collecting data, but rather in creating innovative AI-training processes and architectures:

[B]uilding a truly general reasoning engine will require a more fundamental architectural innovation. What's needed is a way for language models to learn new abstractions that go beyond their training data and have these evolving abstractions influence the model's choices as it explores the space of possible solutions.

We know this is possible because the human brain does it. But it might be a while before OpenAI, DeepMind, or anyone else figures out how to do it in silicon.⁶⁸

Furthermore, it is worth noting that the data most relevant to startups in a given market may not be those held by large incumbent platforms in other markets, but rather data specific to the market in which the startup is active or, even better, to the given problem it is attempting to solve:

As Andres Lerner has argued, if you wanted to start a travel business, the data from Kayak or Priceline would be far more relevant. Or if you wanted to start a ride-sharing business, data from cab companies would be more useful than the broad, market-cross-cutting profiles Google and Facebook have. Consider companies like Uber, Lyft and

⁶⁴ See, e.g., Xiaoliang Dai, et al., *Emu: Enhancing Image Generation Models Using Photogenic Needles in a Haystack*, ARXIV (27 Sep. 2023) at 1, <https://arxiv.labs.arxiv.org/html/2309.15807> (“[S]upervised fine-tuning with a set of surprisingly small but extremely visually appealing images can significantly improve the generation quality.”); see also, Hu Xu, et al., *Demystifying CLIP Data*, ARXIV (28 Sep. 2023), <https://arxiv.org/abs/2309.16671>.

⁶⁵ Lauren Leffer, *New Training Method Helps AI Generalize like People Do*, SCI. AM. (26 Oct. 2023), <https://www.scientificamerican.com/article/new-training-method-helps-ai-generalize-like-people-do> (discussing Brendan M. Lake & Marco Baroni, *Human-Like Systematic Generalization Through a Meta-Learning Neural Network*, 623 NATURE 115 (2023)).

⁶⁶ Timothy B. Lee, *The Real Research Behind the Wild Rumors about OpenAI's Q* Project*, ARS TECHNICA (Dec. 8, 2023), <https://arstechnica.com/ai/2023/12/the-real-research-behind-the-wild-rumors-about-openais-q-project>.

⁶⁷ *Id.*; see also GSM8K, PAPERS WITH CODE (last visited 18 Jan. 2023), available at <https://paperswithcode.com/dataset/gsm8k>; MATH Dataset, GITHUB (last visited 18 Jan. 2024), available at <https://github.com/hendrycks/math>.

⁶⁸ Lee, *supra* note 66.

Sidecar that had no customer data when they began to challenge established cab companies that *did* possess such data. If data were really so significant, they could never have competed successfully. But Uber, Lyft and Sidecar have been able to effectively compete because they built products that users wanted to use—they came up with an idea for a better mousetrap. The data they have accrued came *after* they innovated, entered the market and mounted their successful challenges—not before.⁶⁹

The bottom line is that data is not the be-all and end-all that many in competition circles make it out to be. While data often may confer marginal benefits, there is little sense that these benefits are ultimately decisive.⁷⁰ As a result, incumbent platforms' access to vast numbers of users and troves of data in their primary markets might only marginally affect their competitiveness in AI markets.

A related observation is that firms' capabilities and other features of their products arguably play a more important role than the data they own.⁷¹ Examples of this abound in digital markets. Google overthrew Yahoo in search, despite initially having access to far fewer users and far less data; Google and Apple overcame Microsoft in the smartphone operating system 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,⁷² and TikTok's clever algorithm) appear to have played a far more significant role than initial user and data endowments (or lack thereof).

All of this suggests that the early success of OpenAI likely has more to do with its engineering decisions than with what data it did or did not possess. Going forward, OpenAI and its rivals' ability to offer and monetise compelling stores offering custom versions of their generative-AI technology will arguably play a much larger role than (and contribute to) their ownership of data.⁷³ In other words, the ultimate challenge is arguably to create a valuable platform, of which data ownership is a consequence, but not a cause.

⁶⁹ Geoffrey Manne & Ben Sperry, *Debunking the Myth of a Data Barrier to Entry for Online Services*, TRUTH ON THE MARKET (26 Mar. 2015), <https://truthonthemarket.com/2015/03/26/debunking-the-myth-of-a-data-barrier-to-entry-for-online-services> (citing Andres V. Lerner, *The Role of 'Big Data' in Online Platform Competition* (26 Aug. 2014), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2482780).

⁷⁰ See Catherine Tucker, *Digital Data as an Essential Facility: Control*, CPI ANTITRUST CHRON. (Feb. 2020), at 11 (“[U]ltimately the value of data is not the raw manifestation of the data itself, but the ability of a firm to use this data as an input to insight”).

⁷¹ Or, as John Yun put it, data is only a small component of digital firms' production function. See Yun, *supra* note 28, at 235 (“Second, while no one would seriously dispute that having more data is better than having less, the idea of a data-driven network effect is focused too narrowly on a single factor improving quality. As mentioned in *supra* Section I.A, there are a variety of factors that enter a firm's production function to improve quality”).

⁷² Luxia Le, *The Real Reason Windows Phone Failed Spectacularly*, HISTORY-COMPUTER (8 Aug. 2023), <https://history-computer.com/the-real-reason-windows-phone-failed-spectacularly>.

⁷³ *Introducing the GPT Store*, OPEN AI (Jan. 10, 2024), <https://openai.com/blog/introducing-the-gpt-store>.

It is also important to note that, in those instances where it is valuable, data does not just fall from the sky. Instead, it is through smart business and engineering decisions that firms can generate valuable *information* (which does not necessarily correlate with owning more *data*).

For instance, OpenAI's success with ChatGPT is often attributed to its more efficient algorithms and training models, which arguably have enabled the service to improve more rapidly than its rivals.⁷⁴ Likewise, the ability of firms like Meta and Google to generate valuable data for advertising arguably depends more on design decisions that elicit the right data from users, rather than the raw number of users in their networks.

Put differently, setting up a business so as to extract and organise the right *information* is more important than simply owning vast troves of *data*.⁷⁵ Even in those instances where high-quality data is an essential parameter of competition, it does not follow that having vaster databases or more users on a platform necessarily leads to better information for the platform.

Indeed, if data ownership consistently conferred a significant competitive advantage, these new firms would not be where they are today. This does not, of course, mean that data is worthless. Rather, it means that competition authorities should not assume that the mere possession of data is a dispositive competitive advantage, absent compelling empirical evidence to support such a finding. In this light, the current wave of decisions and competition-policy pronouncements that rely on data-related theories of harm are premature.

IV. Five Key Takeaways: Reconceptualizing the Role of Data in Generative-AI Competition

As we explain above, data network effects are not the source of barriers to entry that they are sometimes made out to be. The picture is far more nuanced. Indeed, as economist Andres Lerner demonstrated almost a decade ago (and the assessment is only truer today):

Although the collection of user data is generally valuable for online providers, *the conclusion that such benefits of user data lead to significant returns to scale and to the entrenchment of dominant online platforms is based on unsupported assumptions.* Although, in theory, control of an “essential” input can lead to the exclusion of rivals, a

⁷⁴ See Michael Schade, *How ChatGPT and Our Language Models are Developed*, OPENAI, <https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-language-models-are-developed>; Sreejani Bhattacharyya, *Interesting Innovations from OpenAI in 2021*, AIM (1 Jan. 2022), <https://analyticsindiamag.com/interesting-innovations-from-openai-in-2021>; Danny Hernandez & Tom B. Brown, *Measuring the Algorithmic Efficiency of Neural Networks*, ARXIV (8 May 2020), <https://arxiv.org/abs/2005.04305>.

⁷⁵ See Yun, *supra* note 28 at 235 (“Even if data is primarily responsible for a platform’s quality improvements, these improvements do not simply materialize with the presence of more data—which differentiates the idea of data-driven network effects from direct network effects. A firm needs to intentionally transform raw, collected data into something that provides analytical insights. This transformation involves costs including those associated with data storage, organization, and analytics, which moves the idea of collecting more data away from a strict network effect to more of a ‘data opportunity.’”).

careful analysis of real-world evidence indicates that such concerns are unwarranted for many online businesses that have been the focus of the “big data” debate.⁷⁶

While data can be an important part of the competitive landscape, incumbents’ data advantages are far less pronounced than today’s policymakers commonly assume. In that respect, five main lessons emerge:

1. Data can be (very) valuable, but beyond a certain threshold, those benefits tend to diminish. In other words, having the *most* data is less important than having *enough*;
2. The ability to generate valuable information does not depend on the number of users or the amount of data a platform has previously acquired;
3. The most important datasets are not always proprietary;
4. Technological advances and platforms’ engineering decisions affect their ability to generate valuable information, and this effect swamps effects stemming from the amount of data they own; and
5. *How* platforms use data is arguably more important than *what* data or *how much* data they own.

These lessons have important ramifications for policy debates over the competitive implications of data in technologically evolving areas.

First, it is not surprising that startups, rather than incumbents, have taken an early lead in generative AI (and in Web 2.0 before it). After all, if data-incumbency advantages are small or even nonexistent, then smaller and more nimble players may have an edge over established tech platforms. This is all the more likely given that, despite significant efforts, the biggest tech platforms were unable to offer compelling generative-AI chatbots and image-generation services before the emergence of ChatGPT, Dall-E, Midjourney, etc.

This failure suggests that, in a process akin to Clayton Christensen’s “innovator’s dilemma”,⁷⁷ something about the incumbent platforms’ existing services and capabilities was holding them back in those markets. Of course, this does not necessarily mean that those same services or capabilities could not become an advantage when the generative-AI market starts addressing issues of monetisation and scale.⁷⁸ But it does mean that assumptions about a firm’s market power based on its possession of data are off the mark.

Another important implication is that, paradoxically, policymakers’ efforts to prevent Web 2.0 platforms from competing freely in generative-AI markets may ultimately backfire and lead to *less*, not more, competition. Indeed, OpenAI is currently acquiring a sizeable lead in generative AI. While

⁷⁶ Lerner, *supra* note 69, at 4-5 (emphasis added).

⁷⁷ See CLAYTON M. CHRISTENSEN, *THE INNOVATOR'S DILEMMA: WHEN NEW TECHNOLOGIES CAUSE GREAT FIRMS TO FAIL* (2013).

⁷⁸ See DAVID J. TEECE, *DYNAMIC CAPABILITIES AND STRATEGIC MANAGEMENT: ORGANIZING FOR INNOVATION AND GROWTH* (2009).

competition authorities might like to think that other startups will emerge and thrive in this space, it is important not to confuse desires with reality. While there currently exists a vibrant AI-startup ecosystem, there is at least a case to be made that the most significant competition for today's AI leaders will come from incumbent Web 2.0 platforms—although nothing is certain at this stage.

Policymakers should beware not to stifle that competition on the misguided assumption that competitive pressure from large incumbents is somehow less valuable to consumers than that which originates from smaller firms. This is particularly relevant in the context of merger control. An acquisition (or an “acqui-hire”) by a “big tech” company does not only, in principle, entail a minor risk to harm competition (it is not a horizontal merger⁷⁹) but could create a stronger competitor to the current market leaders.

Finally, even if there were a competition-related market failure to be addressed in the field of generative AI (which is anything but clear), the remedies under contemplation may do more harm than good. Some of the solutions that have been put forward have highly ambiguous effects on consumer welfare. Scholars have shown that, *e.g.*, mandated data sharing—a solution championed by EU policymakers, among others—may sometimes dampen competition in generative-AI markets.⁸⁰ This is also true of legislation like the General Data Protection Regulation (GDPR), which makes it harder for firms to acquire more data about consumers—assuming such data is, indeed, useful to generative-AI services.⁸¹

In sum, it is a flawed understanding of the economics and practical consequences of large agglomerations of data that lead competition authorities to believe that data-incumbency advantages are likely to harm competition in generative AI markets—or even in the data-intensive Web 2.0 markets that preceded them. Indeed, competition or regulatory intervention to “correct” data barriers and data network and scale effects is liable to do more harm than good.

⁷⁹ Antitrust merger enforcement has long assumed that horizontal mergers are more likely to cause problems for consumers than the latter. *See*: Geoffrey A. Manne, Dirk Auer, Brian Albrecht, Eric Fruits, Daniel J. Gilman, & Lazar Radic, *Comments of the International Center for Law and Economics on the FTC & DOJ Draft Merger Guidelines*, (18 Sep. 2023), <https://laweconcenter.org/resources/comments-of-the-international-center-for-law-and-economics-on-the-ftc-doj-draft-merger-guidelines>.

⁸⁰ *See* Hagiu & Wright, *supra* note 32, at 32 (“We use our dynamic framework to explore how data sharing works: we find that it increases consumer surplus when one firm is sufficiently far ahead of the other by making the laggard more competitive, but it decreases consumer surplus when the firms are sufficiently evenly matched by making firms compete less aggressively, which in our model means subsidizing consumers less.”); *see also* Lerner, *supra* note 69.

⁸¹ *See, e.g.*, Hagiu & Wright, *id.* (“We also use our model to highlight an unintended consequence of privacy policies. If such policies reduce the rate at which firms can extract useful data from consumers, they will tend to increase the incumbent's competitive advantage, reflecting that the entrant has more scope for new learning and so is affected more by such a policy.”); Jian Jia, Ginger Zhe Jin, & Liad Wagman, *The Short-Run Effects of the General Data Protection Regulation on Technology Venture Investment*, 40 *MARKETING SCI.* 593 (2021) (finding GDPR reduced investment in new and emerging technology firms, particularly in data-related ventures); James Campbell, Avi Goldfarb, & Catherine Tucker, *Privacy Regulation and Market Structure*, 24 *J. ECON. & MGMT. STRAT.* 47 (2015) (“Consequently, rather than increasing competition, the nature of transaction costs implied by privacy regulation suggests that privacy regulation may be anti-competitive”).

