

# FROM DATA MYTHS TO DATA REALITY: WHAT GENERATIVE AI CAN TELL US ABOUT COMPETITION POLICY (AND VICE VERSA)



BY GEOFFREY A. MANNE & DIRK AUER<sup>1</sup>



<sup>1</sup> Tirzah Duren is the Vice President of Policy and Research at the American Consumer Institute where her research focuses on the intersection of antitrust, technology, and consumer policy.

# CPI ANTITRUST CHRONICLE

## February 2024

### THEORIES OF HARM BECAUSE OF DATA

By Christian Rusche



### DATA IN THE DIGITAL ECONOMY: THE “ANTI-HERO” OF ANTITRUST OR JUST “BAD REPUTATION?”

By Kristof Zetenyi & Dheeraj Chaudhary



### FROM DATA MYTHS TO DATA REALITY: WHAT GENERATIVE AI CAN TELL US ABOUT COMPETITION POLICY (AND VICE VERSA)

By Geoffrey A. Manne & Dirk Auer



### NOT ALL E-COMMERCE BUSINESSES ARE BORN EQUAL: IMPLICATIONS FOR PRIVACY REGULATION

By Jorge Padilla, Salvatore Piccolo, Albert Riera & Helder Vasconcelos



### THE ECONOMICS OF THE DATA-DRIVEN ECONOMY AND THE DEMAND FOR ANTITRUST

By Dan Ciuriak



### CAN DATA REGULATION RESHAPE THE SHORT-TERM RENTAL MARKET?

By Ginger Zhe Jin, Liad Wagman & Mengyi Zhong



### DATA AT THE FRONTLINE: THE ECONOMIC THEORIES OF DATA-CENTRIC ANTITRUST ENFORCEMENT

By James Mellsop, Veronica Postal & John Scalf



## FROM DATA MYTHS TO DATA REALITY: WHAT GENERATIVE AI CAN TELL US ABOUT COMPETITION POLICY (AND VICE VERSA)

By Geoffrey A. Manne & Dirk Auer

It is often said that so-called “data network effects” entrench incumbents in both the markets where data is collected and adjacent markets. If true, this suggests the largest online platforms — including Alphabet, Meta, Apple, and Amazon — should have a meaningful advantage in the burgeoning markets for generative AI services. And yet the tech giants have, to date, been unable to leverage their vast data troves to outcompete startups like OpenAI and Midjourney. This article suggests there are important lessons about the economics and competitive implications of data to be gleaned from this. As we explain, the rapid emergence of generative AI technology undercuts several core assumptions of today’s digital competition policy debates. Economic findings suggest incumbent data advantages are far less pronounced than today’s policymakers commonly assume. It is thus not surprising that startups 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. Accordingly, policymakers’ efforts to prevent Web 2.0 platforms from competing freely in generative AI markets may backfire by eliminating an important source of competition. In short, competition or regulatory interventions to “correct” alleged data barriers and data network and scale effects are liable to do more harm than good.

Visit [www.competitionpolicyinternational.com](http://www.competitionpolicyinternational.com) for access to these articles and more!

CPI Antitrust Chronicle February 2024

[www.competitionpolicyinternational.com](http://www.competitionpolicyinternational.com)

### Scan to Stay Connected!

Scan or click here to sign up for CPI's FREE daily newsletter.



# I. INTRODUCTION

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.”<sup>2</sup> Similar epithets have been hurled at virtually all large online platforms, including Facebook (Meta), Amazon, and Uber.<sup>3</sup>

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),<sup>4</sup> a shiny new data target has emerged in the form of generative artificial intelligence. 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 people’s conception of what is, and what might be, possible to achieve with generative AI technologies built on massive data sets.

While these services remain in the early stages of mainstream adoption and are in the throes of rapid, unpredictable technological evolution, they nevertheless already appear 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.<sup>5</sup> 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.<sup>6</sup> As Lina Khan, Chair of the FTC, put it: “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”.<sup>7</sup>

In that sense, the response from the competition-policy world is deeply troubling. Instead of engaging in critical self-assessment and adopting an appropriately restrained stance, the enforcement community appears to be chomping 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 deploy existing competition tools rapidly and almost reflexively to address the presumed competitive failures presented by generative AI.<sup>8</sup>

It is increasingly common for competition enforcers to argue that so-called “data network effects” serve not only to entrench incumbents in the markets where that data is collected, but also 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

---

2 Nathan Newman, *Taking on Google’s Monopoly Means Regulating Its Control of User Data*, HUFFINGTON POST (Sep. 24, 2013), [http://www.huffingtonpost.com/nathan-newman/taking-on-googlesmonopol\\_b\\_3980799.html](http://www.huffingtonpost.com/nathan-newman/taking-on-googlesmonopol_b_3980799.html).

3 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. (Oct. 1, 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.”).

4 See generally Abigail Slater, *Why “Big Data” Is a Big Deal*, THE REG. REV. (Nov. 6, 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.”).

5 See e.g. Press Release, European Commission, *Commission Launches Calls for Contributions on Competition in Virtual Worlds and Generative AI* (Jan. 9, 2024), [https://ec.europa.eu/commission/presscorner/detail/en/IP\\_24\\_85](https://ec.europa.eu/commission/presscorner/detail/en/IP_24_85); Krysten Crawford, *FTC’s Lina Khan warns Big Tech over AI*, SIEPR (Nov. 3, 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).

6 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, *Are New EU Data Market Regulations Coherent and Efficient?*, Bruegel Working Paper 21/23 (2023), available at <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 (Feb. 24, 2021), <https://www.oecd-forum.org/posts/competitive-dysfunction-why-competition-law-is-failing-in-a-digital-world>.

7 Rana Foroohar, *The Great US-Europe Antitrust Divide*, FT (Feb. 5, 2024), <https://www.ft.com/content/065a2f93-dc1e-410c-ba9d-73c930cedc14>.

8 See e.g. Press Release, European Commission, *supra* note 5.

access to data to extend their dominance into these new markets.<sup>9</sup> 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?”<sup>10</sup> Unsurprisingly, the U.S. Federal Trade Commission (“FTC”) has 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 control over these inputs to unlawfully entrench their market positions in AI and related markets, including digital content markets.<sup>11</sup>

Against this backdrop, 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 recognized that data is an essential input for generative AI.<sup>12</sup> 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 have routinely made headlines.<sup>13</sup> Apple and Amazon also have vast experience with AI assistants, and all of these firms use AI technology throughout their platforms.<sup>14</sup>

Contrary to what one might expect, however, the tech giants have, to date, been unable to leverage their vast data troves to outcompete startups like OpenAI and Midjourney. At the time of writing, OpenAI’s ChatGPT appears to be, by far, the most successful chatbot<sup>15</sup>, despite the fact that large tech platforms arguably have access to far more (and more up-to-date) data.

This article suggests there are important lessons to be learned from the current technological moment, 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 — the rueful after-effects of the purported failure of 20<sup>th</sup>-century antitrust to address the allegedly manifest harms of 21<sup>st</sup>-century technology. These include the notions that data advantages constitute barriers to entry and can be leveraged to project 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.

---

9 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 (Jan. 15, 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.”).

10 Press Release, European Commission, *supra* note 5.

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

12 See, e.g. Joe Caserta, Holger Harreis, Kayvaun Rowshankish, Nikhil Srinidhi, and Asin Tavakoli, *The data dividend: Fueling generative AI*, MCKINSEY DIGITAL (Sept. 15, 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.”).

13 See e.g. Tim Keary, *Google DeepMind’s Achievements and Breakthroughs in AI Research*, TECHOPEDIA (Aug. 11, 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 (Dec. 14, 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 (Nov. 30, 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 Jan. 18, 2023).

14 See e.g. Jennifer Allen, *10 years of Siri: the history of Apple’s voice assistant*, TECH RADAR (Oct. 4, 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 (Nov. 20, 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 (July 19, 2019), <https://www.forbes.com/sites/cognitive-world/2019/07/19/the-twenty-year-history-of-ai-at-amazon/?sh=1734bcb268d0>.

15 See *infra* Section III.

## II. DATA NETWORK EFFECTS THEORY AND ENFORCEMENT

Proponents of tougher interventions 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).<sup>16</sup> 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.”<sup>17</sup> This self-reinforcing cycle purportedly leads to market domination by a single firm. Thus, for Google, for example, it is argued that its “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*.”<sup>18</sup>

Right off the bat, it is important to note the conceptual problem of these claims. Because 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. Of course, it would be a curious approach to antitrust if this were treated as a problem, as it would imply that firms should under-compete — should forego consumer-welfare enhancements—in order to bring about a greater number of firms in a given market simply for its own sake.<sup>19</sup>

Meanwhile, actual economic studies of data network effects are few and far between, with scant empirical evidence to support the theory.<sup>20</sup> Andrei Hagiu and Julian Wright’s theoretical paper offers perhaps the most comprehensive treatment of the topic.<sup>21</sup> The authors ultimately conclude that data network effects can be of different magnitudes and have varying effects on firms’ incumbency advantage.<sup>22</sup> 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.”<sup>23</sup>

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.”<sup>24</sup>

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

---

16 See e.g. Cédric Argenton & Jens Prüfer, *Search Engine Competition with Network Externalities*, 8 J. COMP. L. & ECON. 73, 74 (2012); Mark A. Lemley & Matthew Wansley, *Coopting Disruption* (February 1, 2024), <https://ssrn.com/abstract=4713845>.

17 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., Nov. 11, 2020) at 233, available at [https://gaidigitalreport.com/2020/08/25/big-data-and-barriers-to-entry/#\\_ftnref50](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, available at <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 (Feb. 6, 2024), <https://www.siliconrepublic.com/enterprise/data-ai-aggregation-laws-regulation-big-tech-dominance-competition-antitrust-imd>.

18 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.”).

19 See also Yun, *supra* note 17 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.”).

20 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).

21 Andrei Hagiu & Julian Wright, *Data-Enabled Learning, Network Effects, and Competitive Advantage*, 54 RAND J. ECON. 638 (2023) (final preprint available at <https://andrei-hagiu.com/wp-content/uploads/2022/08/Data-enabled-learning-Final-RAND-Article.pdf>).

22 *Id.* at 2. 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....”

23 *Id.*

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

market. For instance, a search engine could be able to predict new trends from consumer searches and therefore face less uncertainty in product design.<sup>25</sup>

This possibility is also implicit in the paper by Hagiu and Wright.<sup>26</sup> 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).

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 adjacent ones (be it via merger or organic growth). As we explain below, however, there is ultimately little evidence to support such claims.

Policymakers, however, have largely been receptive to these limited theoretical findings, basing multiple decisions on these theories, often with little consideration of the caveats that accompany them.<sup>27</sup> Indeed, it is remarkable that, in the Furman Report's section on "[t]he data advantage for incumbents," only two empirical economic studies are cited, and they offer *directly contradictory conclusions* with respect to the question of the strength of data advantages.<sup>28</sup> Nevertheless, the Furman Report concludes that data "may confer a form of unmatched advantage on the incumbent business, making successful rivalry less likely,"<sup>29</sup> and adopts without reservation "convincing" evidence from non-economists with apparently no empirical basis.<sup>30</sup>

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.<sup>31</sup>

As a result, the Commission cleared the merger on the condition that Google refrain from using data from Fitbit devices for its advertising platform.<sup>32</sup> The Commission will likely focus on similar issues during its ongoing investigation into Microsoft's investment into OpenAI.<sup>33</sup>

Along similar lines, the FTC's complaint to enjoin Meta's purchase of a virtual-reality (VR) fitness app called "Within" relied, among other things, on 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."<sup>34</sup>

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

26 See Hagiu & Wright, *supra* note 21.

27 For a summary of these limitations, see generally Catherine Tucker, *Network Effects and Market Power: What Have We Learned in the Last Decade?*, ANTITRUST (Spring 2018) at 72, available at <https://sites.bu.edu/tpri/files/2018/07/tucker-network-effects-antitrust2018.pdf>. See also Manne & Auer, *supra* note 20, at 1330.

28 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](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/785547/unlocking_digital_competition_furman_review_web.pdf).

29 *Id.* at 34.

30 *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 does maintain 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.

31 Case COMP/M.9660 — Google/Fitbit, Commission Decision (Dec. 17, 2020) (Summary at O.J. (C 194) 7), available at [https://ec.europa.eu/competition/mergers/cases1/202120/m9660\\_3314\\_3.pdf](https://ec.europa.eu/competition/mergers/cases1/202120/m9660_3314_3.pdf) at 455.

32 *Id.* at 896.

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

34 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](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).

The U.S. Department of Justice’s twin cases against Google also raise data leveraging and data barriers to entry. The agency’s AdTech complaint that “Google intentionally exploited its massive trove of user data to further entrench its monopoly across the digital advertising industry.”<sup>35</sup> Similarly, in its 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.<sup>36</sup>

Finally, the merger guidelines published by several competition enforcers cite the acquisition of data as a potential source of competitive concerns. For instance, the FTC and DOJ’s newly published 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.”<sup>37</sup> Likewise, the UK Competition and Markets Authority (“CMA”) warns against incumbents acquiring firms in order to obtain their data and foreclose other rivals:

Incentive to foreclose rivals . . .

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....<sup>38</sup>

In short, competition authorities around the globe are taking an aggressive stance on data network effects. Among the ways this has manifested is in basing enforcement decisions 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 canvassed in the previous section, it seems reasonable to assume that firms such as Google, Meta, and Amazon would 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.”<sup>39</sup>

At the time of writing, 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).<sup>40</sup> 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 the previous record holder, TikTok.<sup>41</sup> Based on Google Trends data, ChatGPT is nine times more popular than Google’s own Bard service worldwide, and 14

---

35 Amended Complaint (D.D.C.), *supra* note 4, at ¶37.

36 Amended Complaint (E.D. Va.), *supra* note 4, at ¶8.

37 US Dep’t of Justice & Fed. Trade Comm’n, MERGER GUIDELINES (2023) at 25, [https://www.ftc.gov/system/files/ftc\\_gov/pdf/2023\\_merger\\_guidelines\\_final\\_12.18.2023.pdf](https://www.ftc.gov/system/files/ftc_gov/pdf/2023_merger_guidelines_final_12.18.2023.pdf).

38 Competition and Mkts. Auth., MERGER ASSESSMENT GUIDELINES (2021) at ¶7.19(e), [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1051823/MAGs\\_for\\_publication\\_2021\\_--\\_.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1051823/MAGs_for_publication_2021_--_.pdf).

39 Furman Report, *supra* note 28, at ¶4.

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

41 See Krystal Hu, *ChatGPT Sets Record for Fastest-Growing User Base*, REUTERS (Feb. 2, 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 (Feb. 7, 2023), <https://www.appeconomyinsights.com/p/google-the-ai-race-is-on>.

times more popular in the U.S.<sup>42</sup> In April 2023, ChatGPT reportedly registered 206.7 million unique visitors, compared to 19.5 million for Google's Bard.<sup>43</sup> In short, at the time of writing, ChatGPT appears to be the most popular chatbot. And, so far, the entry of large players such as Google Bard or Meta AI appear to have had little effect on its market position.<sup>44</sup>

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.<sup>45</sup> 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.<sup>46</sup>

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, some observations concerning the role and value of data in digital markets would appear to be relevant.

A first important observation is that empirical studies suggest 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 puts 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."<sup>47</sup>

Likewise, following a survey of the empirical literature on this topic, Geoffrey Manne & 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 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.<sup>48</sup>

In other words, being the firm with the *most* data appears to be far less important than having *enough* data, and this lower bar may be accessible to far more firms than one might initially think possible.

And obtaining enough data could become even easier — 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<sup>49</sup> — or may even outperform real-world data.<sup>50</sup> As Thibault Schrepel and Alex Pentland point out, "advances 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."<sup>51</sup>

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 effect. In fact, there is some evidence that excessive data impedes a service's ability

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

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

44 See Press Release, Meta, *Introducing New AI Experiences Across Our Family of Apps and Devices* (Sept. 27, 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/>.

45 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 (Oct. 13, 2023), <https://approachableai.com/midjourney-statistics/>.

46 See Hema Budaraju, *New Ways to Get Inspired with Generative AI in Search*, GOOGLE KEYWORD BLOG (Oct. 12, 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/>.

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

48 Manne & Auer, *supra* note 20, at 1345.

49 See e.g. Stefanie Koperniak, *Artificial Data Give the Same Results as Real Data—Without Compromising Privacy*, MIT NEWS (Mar. 3, 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.")

50 See e.g. Rachel Gordon, *Synthetic Imagery Sets New Bar in AI Training Efficiency*, MIT NEWS (Nov. 20, 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.")

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



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.”<sup>52</sup>

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, but is surrounded by copious other image data, the user may end up with an inordinately noisy result. By contrast, a model trained with a better method on fewer, more-carefully selected images, could readily yield far superior results.<sup>53</sup> 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.<sup>54</sup>

Current efforts are thus focused on improving the mathematical and logical reasoning of large language models (“LLMs”), rather than maximizing training datasets.<sup>55</sup> Two points stand out. The first is that firms like OpenAI rely largely on publicly available datasets — such as GSM8K — to train their LLMs.<sup>56</sup> 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.<sup>57</sup>

Furthermore, it is worth noting that the data most relevant to startups operating in a given market may not be those data 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 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.<sup>58</sup>

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

53 See e.g. Xiaoliang Dai, et al., *Emu: Enhancing Image Generation Models Using Photogenic Needles in a Haystack*, ArXiv (Sep. 27, 2023) at 1, <https://ar5iv.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 (Sep. 28, 2023), <https://arxiv.org/abs/2309.16671>.

54 Lauren Leffer, *New Training Method Helps AI Generalize like People Do*, Sci. Am. (Oct. 26, 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)).

55 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/>.

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

57 Lee, *supra* note 55.

58 Geoffrey Manne & Ben Sperry, *Debunking the Myth of a Data Barrier to Entry for Online Services*, TRUTH ON THE MARKET (Mar. 26, 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* (Aug. 26, 2014), available at [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2482780](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2482780)).

The bottom line is that data is not the be-all and end-all that many in competition circles rather casually make it out to be.<sup>59</sup> While data may often confer marginal benefits, there is little sense these are ultimately decisive.<sup>60</sup> As a result, incumbent platforms' access to vast numbers of users and data in their primary markets might only marginally affect their AI competitiveness.

A related observation is that firms' capabilities and other features of their products arguably play a more important role than the data they own.<sup>61</sup> Examples of this abound in digital markets. 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,<sup>62</sup> and TikTok's clever algorithm) appear to have played a far greater 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 the data it did (or did not) own. And going forward, OpenAI and its rivals' ability to offer and monetize 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.<sup>63</sup> In other words, the ultimate challenge is arguably to create a valuable platform, of which data ownership is a consequence, but not a cause.

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 owing 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.<sup>64</sup> 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 generate the right *information* is more important than simply owning vast troves of *data*.<sup>65</sup> 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.

Given what precedes, it seems clear that OpenAI and other generative AI startups' early success, as well as their chances of prevailing in the future, hinge on a far broader range of factors than the mere ownership of data. Indeed, if data ownership consistently conferred a significant competitive advantage, these new firms would not be where they are today. This does not mean that data is worthless, of course. Rather, it means that competition authorities should not assume that merely possessing 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.

---

59 See e.g., Lemley & Wansley, *supra* note 18, at 22 ("Incumbents have all that information. It would be difficult for a new entrant to acquire similar datasets independently. . .").

60 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.").

61 Or, as John Yun puts it, data is only a small component of digital firms' production function. See Yun, *supra* note 17, 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.").

62 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/>.

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

64 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 (Jan. 1, 2022), <https://analyticsindiamag.com/interesting-innovations-from-openai-in-2021/>; Danny Hernandez & Tom B. Brown, *Measuring the Algorithmic Efficiency of Neural Networks*, ARXIV (May 8, 2020), available at <https://arxiv.org/abs/2005.04305>.

65 See Yun, *supra* note 17 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.'").

## 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; rather, 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 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.<sup>66</sup>

While data can be an important part of the competitive landscape, incumbent 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 past a certain threshold, the 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 the effect of 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 competition-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 Christensen’s Innovator’s Dilemma,<sup>67</sup> something about their existing services and capabilities was holding them back in those markets. Of course, this does not necessarily mean that those same services/capabilities could not become an advantage when the generative AI market starts addressing issues of monetization and scale.<sup>68</sup> But it does mean that assumptions of 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 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. For, while there is 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.

Finally, even if there were a competition-related market failure to be addressed (which is anything but clear) in the field of generative AI, it is unclear that contemplated remedies would do more good than harm. Some of the solutions that have been put forward have highly ambiguous effects on consumer welfare. Scholars have shown that mandated data sharing — a solution championed by EU policymakers, among

---

<sup>66</sup> Lerner, *supra* note 58, at 4-5 (emphasis added).

<sup>67</sup> See CLAYTON M. CHRISTENSEN, *THE INNOVATOR’S DILEMMA: WHEN NEW TECHNOLOGIES CAUSE GREAT FIRMS TO FAIL* (2013).

<sup>68</sup> See DAVID J. TEECE, *DYNAMIC CAPABILITIES AND STRATEGIC MANAGEMENT: ORGANIZING FOR INNOVATION AND GROWTH* (2009).

others — may sometimes dampen competition in generative AI markets.<sup>69</sup> This is also true of legislation like the GDPR that make it harder for firms to acquire more data about consumers — assuming such data is, indeed, useful to generative AI services.<sup>70</sup>

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.

---

69 See Hagiu and Wright, *supra* note 21, at 4 (“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 58.

70 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.”).

## CPI Subscriptions

CPI reaches more than 35,000 readers in over 150 countries every day. Our online library houses over 23,000 papers, articles and interviews.

Visit [competitionpolicyinternational.com](http://competitionpolicyinternational.com) today to see our available plans and join CPI's global community of antitrust experts.

