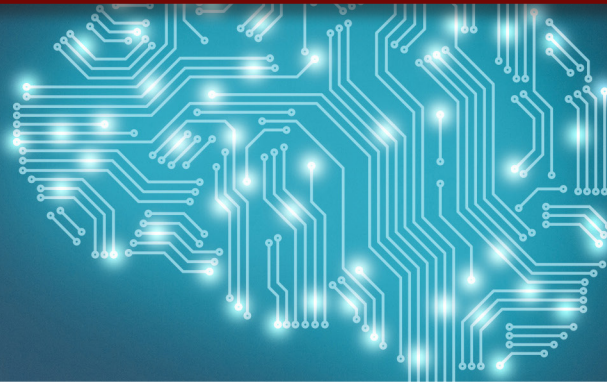


ARTIFICIAL INTELLIGENCE, INCENTIVES TO INNOVATE, AND COMPETITION POLICY



BY SAMUEL HIMEL & ROBERT SEAMANS¹



I. INTRODUCTION

The role of artificial intelligence (“AI”) in our economy and our society is growing rapidly and is affecting a variety of business services — including consumer advertising, financial advice, and insurance — as well as government functions such as law enforcement and criminal sentencing.² Large datasets are a critical input for firms that want to create or use AI systems. Even the best AI algorithms are useless without an underlying large-scale dataset, because large datasets are needed for the initial training and fine-tuning of these algorithms.

As a result, the growing importance of data has been highlighted recently by *The Economist* (among others), which has likened its value today to that held by oil for much of the past century.³ Antitrust enforcement officials have already recognized that challenges may arise when large incumbent firms control the vast majority of such data. For example, FTC Commissioner Terrell McSweeney has noted, “It may be that an incumbent has significant advantages over new entrants when a firm has a database that would be difficult, costly, or time consuming for a new firm to match or replicate.”⁴

If firms face high barriers to accessing such datasets, then they may opt not to enter a market that requires large datasets as inputs, leading to less competition in that market. Both startups and existing firms may forgo entry because of this difficulty, and so competition would decline in both new and established markets. In general, a lack of competition hurts consumers, in some cases via higher prices and in other cases via a reduction in the number of improved features, new products or other innovations.⁵ While there

1 Robert Seamans is Associate Professor of Management and Operations in the Stern School of Business at New York University, where Samuel Himel is a second-year student in the School of Law. The authors are grateful for comments and advice from Hanna Halaburda, Gabriel Scheffler, and Christopher Sprigman.

2 CAMPOLO ET AL., AI NOW 2017 REPORT 3 (Selbst & Barocas, eds., 2017).

3 *Fuel of the Future. Data is Giving Rise to a New Economy. How Is It Shaping Up?*, THE ECONOMIST, May 6, 2017, <https://www.economist.com/news/briefing/21721634-how-it-shaping-up-data-giving-rise-new-economy>.

4 Commissioner Terrell McSweeney, Opening Remarks for a Panel Discussion, “Why Regulate Online Platforms?: Transparency, Fairness, Competition, or Innovation?” at the CRA Conference in Brussels, Belgium, at 5 (Dec. 9, 2015), https://www.ftc.gov/system/files/documents/public_statements/903953/mcsweeney_-_cra_conference_remarks_9-12-15.pdf.

5 COUNCIL OF ECON. ADVISERS, BENEFITS OF COMPETITION AND INDICATORS OF MARKET POWER 2, (May 2016), https://obamawhitehouse.archives.gov/sites/default/files/page/files/20160502_competition_issue_brief_updated_cea.pdf [hereinafter “CEA Report”].

has historically been some debate as to whether a decline in competition leads to higher or lower levels of innovation — a debate we briefly recap below — current evidence suggests we are in an era of low competition in a number of economic sectors, and hence more competition is likely to lead to more innovation and ultimately be beneficial to consumers.⁶ Moreover, even if one were to assume that the evidence is equivocal regarding the overall relationship between competition and innovation, increased competition is likely to improve innovation in those key markets affected by AI development.

To this end, new enforcement policies and regulatory strategies may be needed to ensure that both incumbent and potential entrant firms have access to the datasets they need to innovate in the AI domain. In the latter scenario, this effort will enable entry — what is sometimes referred to as competition *for* the market — and in the former one, it will increase the level of competition among incumbent firms *in* the market. In both settings, we expect the associated broader access to data to foster more innovation.

In the rest of this piece, we discuss how antitrust enforcers might address these issues, while highlighting two challenges with such approaches: the length of time necessary for resultant remedies to take effect, and the difficulty of handling non-price transactions, which have become more frequent due to an increase in the prevalence of firms operating in two-sided markets. We also describe several novel policy and regulatory solutions for potential future consideration, including provisions that would institute temporary data monopolies, data portability regimes, the use of trusted third parties, and blockchain-enabled technological solutions. We discuss how several of these approaches are complementary to each other. We also contrast the ways in which each approach differs in its ability to safeguard consumer data and privacy, incentivize incumbent behavior, enable competition between incumbent firms, and facilitate entry by startup firms.

II. THE GROWING IMPORTANCE OF AI TO THE ECONOMY AND SOCIETY

The news media regularly provides updates on advances in AI, including its ability to defeat humans in complex games such as chess, go and poker. As tracked by the Electronic Frontier Foundation, AI systems are rapidly approaching or surpassing human expertise at a number of other tasks.⁷ A number of firms are now using AI to enhance existing products or to offer brand new ones. For example, AI-enabled virtual assistants like Apple's Siri now come bundled in an iPhone, and Amazon's Alexa is available for purchase at Whole Foods Markets.

By many measures, investment in AI has also rapidly increased. Much of this investment appears to be done by established firms. The McKinsey Global Institute ("MGI") estimates that established firms spent between \$18 and \$27 billion on internal corporate investment in AI-related projects in 2016.⁸ Such firms also spend money on AI-related investments in the form of acquisitions. Facebook, Google, Amazon and Apple have bought up hundreds of innovative startups over the past decade, including ones that focus on AI or AI-related technologies.⁹ MGI also notes that established firms spent \$2 to \$3 billion on AI-related M&A in 2016 alone.¹⁰ While less in dollar value, investment in AI-related startups has also been increasing — our analysis of Crunchbase data indicates an increase in such funding that begins in around 2012 and then accelerates sharply in 2014 (Figure below). This observation corroborates a recent report by McKinsey Global Institute that venture capital investment in AI startups grew by 40 percent between 2013 and 2016.¹¹

6 Id.; this conclusion follows from the notion that the relationship between competition and innovation may appear uncertain because innovation is concave in competition. See Aghion et al., *Competition and Innovation: An Inverted-U Relationship*, 120 THE QUARTERLY J. OF ECON. 701, 701-02 (2005).

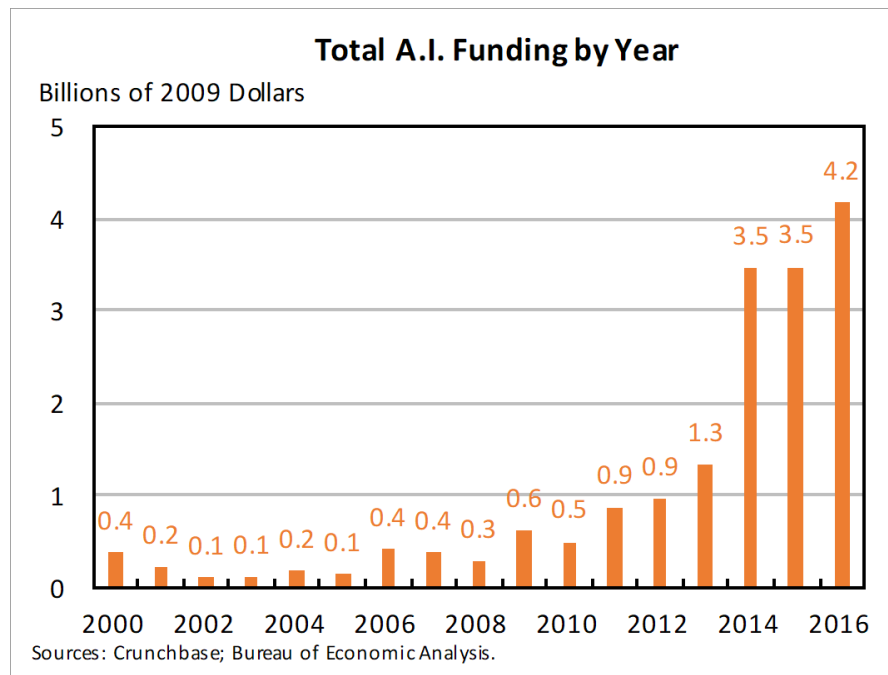
7 *AI Progress Measurement*, ELECTRONIC FRONTIER FOUND., <https://www EFF.ORG/AI/METRICS>.

8 BUGHIN ET AL., ARTIFICIAL INTELLIGENCE: THE NEXT DIGITAL FRONTIER? 10 (McKinsey Global Inst., June 2017), <http://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/how-artificial-intelligence-can-deliver-real-value-to-companies> [hereinafter "MGI Report"].

9 Tech Platforms Weekly: A Closer Look at Amazon's Conduct in the Book Market; More Claims of Search Bias; Facebook, Apple, and Net Neutrality Updates; The Myspace Myth, THE CAPITOL FORUM (Jan. 20, 2017) <http://thecapitolforum.cmail2.com/t/ViewEmail/j/91CFEB1924D56C52/45A74A929A973E10E663AB054A538FBA>.

10 MGI Report, *supra* note 8, at 10.

11 Id.



The important role of data as an input to this growing market for AI — combined with the potential economic implications of the fact that a small number of firms control large datasets on consumers and their prior purchasing behavior — has given rise to calls to update existing antitrust frameworks to address these issues. The challenge for litigation, policy and regulation in this area is the need to advance multiple objectives, particularly the safeguarding of consumer data and consumer privacy while incentivizing firms to innovate and compete for consumers' benefit. In the remaining sections, we first describe the relationship between competition and innovation as it pertains to data and AI. We then highlight how existing litigation strategies could be applied to achieve these goals while protecting consumers. Finally, we discuss additional, more novel policy and regulatory solutions that could complement an enforcement approach in the future.

III. THE RELATIONSHIP BETWEEN COMPETITION AND INNOVATION

Broadly speaking, economists and legal academics have not settled on a consensus regarding an overarching relationship between competition and innovation that holds true economy-wide.¹² The evidence has alternately suggested that a decline in competition in a particular market will increase, decrease or have no effect on innovation levels. This uncertainty is often known as the Arrow-Schumpeter debate. Economist Kenneth Arrow is seen as espousing the view that competition is more conducive to innovation than monopoly, while economist Joseph Schumpeter championed the notion that monopolies promote innovation and naturally give rise to R&D investments.¹³ While the federal antitrust enforcement authorities have in their *Horizontal Merger Guidelines* indicated qualified support for the theory that a reduction in competition may be harmful to innovation, the possibility of harming innovation alone is typically insufficient to justify government antitrust intervention.¹⁴

Nevertheless, a variety of considerations suggest that when it comes to AI, policymakers should treat the relationship between competition and innovation as more closely representing Arrow's view than Schumpeter's. As an initial matter, recent empirical evidence points to a multi-sector decline in competition and an increase in concentration that may be associated with a variety of economic trends,

¹² See, e.g. Baker, *Beyond Schumpeter vs. Arrow: How Antitrust Fosters Innovation*, 74 ANTITRUST L.J. 575, 577-88 (2007); Lao, *Ideology Matters in the Antitrust Debate*, 74 ANTITRUST L.J. 649, 659-62 (2014); Katz & Shelanski, *"Schumpeterian" Competition and Antitrust Policy in High-Tech Markets*, 14 COMPETITION 47 (2005).

¹³ Baker, *supra* note 12, at 577; Clement, *Creative Disruption*, FED. RES. BANK OF MINNEAPOLIS (Sep. 1, 2008), <https://www.minneapolisfed.org/publications/the-region/creative-disruption>.

¹⁴ U.S. Dep't of Justice & Fed. Trade Comm'n, *Horizontal Merger Guidelines* 23-24 (2010), available at: www.ftc.gov/sites/default/files/attachments/merger-review/100819hmg.pdf [hereinafter "Merger Guidelines"].

such as reduced firm dynamism, increased firm age, decreased labor mobility and lower total factor productivity growth.¹⁵ More specific to AI, there may be reason to believe that monopolists are less likely to adopt highly disruptive innovations, such as those that give rise to entirely new markets, in order to retain their position.¹⁶ In such settings, competition would thus be more conducive to innovation than monopoly.

Moreover, given the current state of AI technologies, we argue that the most relevant market for regulators and enforcers to consider at the present time from a competition policy perspective is the market for data inputs to AI research and development. Competition in this market takes place “in the market” (i.e. among incumbents who are trying to maintain a competitive advantage, not only for AI inputs but also for the broader array of transactions they conduct that require such data), as well as “for the market” (i.e. among both entrants and incumbents who will be trying to succeed in future markets for AI technologies).¹⁷ Any set of solutions geared towards removing barriers to AI development would focus on bolstering both forms of competition. As noted above, absent competition, incumbents have less incentive to adopt highly disruptive innovations, while entrants to this particular market by definition cannot innovate without access to datasets as inputs to production. Our proposed solutions aim to address competition for large datasets both among incumbent firms and entrants who wish to access such data. These solutions attempt to account for the possibility that the prospect of obtaining a dataset monopoly may provide a key private incentive for investment in the kinds of platforms that generate the large datasets on which AI development will depend.

IV. POTENTIAL LITIGATION STRATEGIES

Antitrust litigation may provide one avenue by which to mitigate deleterious effects on innovation that result from a lack of competition among firms that exchange or collect large quantities of data. Such litigation would aim to use a range of legal claims sounding in antitrust law to increase competition among these firms. That is to say, it would not necessarily focus on legal theories that mention AI exclusively or explicitly. Nevertheless, while this strategy could potentially benefit consumers and firms in a wide range of scenarios — including those far removed from the AI context — its ultimate goal would be to prevent datasets that are crucial for the development of new technologies like AI from being controlled by only a handful of firms.

The downsides to using litigation as a tool for increasing innovation in AI are that it is inflexible, expensive, and requires a substantial amount of time to bring about results. Regardless of whether a private party or government enforcer brings a case, litigation may take too long to benefit the innovative process meaningfully. While the key era in AI development is rapidly approaching, it could take years or even decades for these cases to conclude, especially if they involve complicated industries or especially novel legal theories.

A key aspect of the online markets in which these firms operate is the lack of an explicit or clearly measurable price in many cases, owing to the two- or multi-sided nature of an increasing number of these firms’ business models. Such firms serve distinct groups of customers on each side of their markets, thus facilitating direct interaction among them. For example, Google connects web searchers to advertisers and eBay connects end-customers to sellers. A key feature of these markets is the indirect network effect between these two sides, which may be asymmetric in some cases.¹⁸ Thus, the greater the number of customers there are on one side of the market (e.g. web searchers), the more customers there will be who will be willing to pay on the other side of the market (e.g. advertisers). This feature can lead firms to set prices very low, even at zero, on one side of the market, so as to increase the number of customers on that side, because this will drive revenue even higher on the other, “paying” side of the market.¹⁹ Such low or “zero” prices can be observed

15 See Furman & Orszag, Presentation at “A Just Society” Centennial Event in Honor of Joseph Stiglitz at Columbia University: A Firm-Level Perspective on the Role of Rents in the Rise in Inequality 3-6, 9-13 (Oct. 16, 2015), https://obamawhitehouse.archives.gov/sites/default/files/page/files/20151016_firm_level_perspective_on_role_of_rents_in_inequality.pdf; CEA Report, supra note 5, at 2; *A Lapse in Concentration*, THE ECONOMIST, Sep. 29, 2016; Joshua Wright and Matt Stoller, *Should the Government Bring Back Trust-Busting?*, NY TIMES, Nov. 14, 2016; Asher Schechter, *Economists: “Totality of Evidence” Underscores Concentration Problem in the U.S.*, PROMARKET (Mar. 31, 2017), <https://promarket.org/economists-totality-evidence-underscores-concentration-problem-u-s/>.

16 Clement, supra note 13, at 8 (summarizing the conditions under which the Gilbert-Newbery model does not produce results favoring the Schumpeterian view).

17 Evans & Schmalensee, *Some Economic Aspects of Antitrust Analysis in Dynamically Competitive Industries*, in INNOVATION POLICY AND THE ECONOMY, VOLUME 2 1, 1 (Jaffe, Lerner & Stern, eds., 2002).

18 See Parker & Van Alstyne, *Two Sided Network Effects: A Theory of Information Product Design*, 51 MGMT. SCI. 1494, 1495-96 (2005); Schmalensee & Evans, *The Industrial Organization of Markets with Two-Sided Platforms*, 3 COMPETITION POL’Y INT’L 151, 151-56 (2007); Seamans & Zhu, *Responses to Entry in Multi-Sided Markets: The Impact of Craigslist on Local Newspapers*, 60 MGMT. SCI. 476 (2014).

19 See Rochet & Tirole, *Platform Competition in Two-Sided Markets*, 1 J. OF THE EUROPEAN ECON. ASS’N 990, 1015 (2003); Rochet & Tirole, *Two-Sided Markets: A Progress Report*, 35 THE RAND J. OF ECON. 645, 659 (2006).

on many consumer-facing digital platforms; for example, the price paid to Google for each search result is zero, as is the price paid to Twitter for joining and using its social network. However, even though the price paid by an end-user to Google or Twitter may be zero, the end-user may still “pay” by giving personal data to these companies or via time spent viewing and scrolling through lower quality content and advertisements.²⁰

The fact that the market price is zero gives rise to a substantial obstacle for government authorities and private parties seeking to bring antitrust claims. Although evidence of non-price effects tends to get some degree of credit in U.S. antitrust law from the enforcement authorities and from the courts,²¹ Stucke and Grunes note that “the agencies’ merger review has migrated towards assessing what is measurable.”²² Firms in multisided markets that lack clear prices for some of their transactions may thus make for challenging antitrust litigation targets.²³

The response to this challenge from commentators who favor increased antitrust scrutiny of big technology firms has been to suggest that government agencies and the courts move beyond just considering price-based harm to consumers and increase their focus on non-price effects, including features of product variety and quality such as privacy.²⁴ This recommendation is broadly consistent with calls for courts and enforcers to consider more than just “consumer welfare” (a broad way of construing the current standard).²⁵ Such a move would in some ways amount to a return to the pre-Chicago School era in antitrust thinking when a greater variety of firm arrangements, behaviors and market effects received antitrust scrutiny.²⁶

This solution, however, should only comprise one part of a more comprehensive legal strategy to increase competition among big technology firms. It could be complemented by more conventional, conduct-based litigation theories that would treat a large dataset as a product or service in an intermediate market.²⁷ Harm on the basis of price could also be alleged by analyzing multisided platform markets in unconventional ways.

Finally, focusing on one of the key non-price parameters on which some observers have suggested bringing cases — namely, privacy — would not necessarily lead to improvements in the availability of data for AI innovators, and indeed it could even have the opposite effect.²⁸ While it may improve consumer protection commitments made by incumbents in the realm of privacy, antitrust litigation brought on a theory of privacy harm may even restrict the ability of non-incumbent innovators to access platform companies’ large datasets. Remedies for privacy harm would likely expand privacy protections, thus necessarily barring individuals and other firms from accessing personal data — the very AI inputs that innovators require.²⁹

20 While two-sided markets appear to be more prevalent now, they are not new, nor are the non-price effects we are describing. For example, radio and TV are clear examples of two-sided markets. In the latter two cases, end users did not traditionally pay a price directly to the radio or TV station but did need to “pay” via time spent watching or listening to ads.

21 Such non-price effects include product quality and dynamic effects on the incentives for innovation. See, e.g. Merger Guidelines, *supra* note 14, at 23; Khan, *Amazon’s Antitrust Paradox*, 126 *YALE L.J.* 710, 721-22 (2017); STUCKE & GRUNES, *BIG DATA AND COMPETITION POLICY* 114-115 (2016).

22 Stucke & Grunes, *supra* note 21, at 107.

23 See generally *id.* at 69-104.

24 See, e.g. Khan, *supra* note 21, at 721-22.

25 *Id.* at 716.

26 See generally Rubinfeld, *On the Foundations of Antitrust Law and Economics*, in *HOW THE CHICAGO SCHOOL OVERSHOT THE MARK, THE EFFECT OF CONSERVATIVE ECONOMIC ANALYSIS ON U.S. ANTITRUST* 51, 52-56 (Pitofsky ed. 2008) (summarizing how antitrust thinking has evolved over the last half century).

27 See, e.g. PATTERSON, *ANTITRUST LAW IN THE NEW ECONOMY* 15-16 (2017).

28 See generally *id.* at 163-181 (arguing for the treatment of privacy as a “good” and positing that competition on privacy could be evaluated by antitrust authorities); Stucke & Grunes, *supra* note 21, at 141-154 (highlighting the difficulty that competition enforcers face in balancing privacy concerns with other issues).

29 To be sure, this litigation strategy may still be worth pursuing if one believes that the benefits of consumer privacy protection outweigh the forgone benefits of AI innovation. See Calo, *Artificial Intelligence Policy: A Primer and Roadmap* 19, available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3015350 (“[P]rivacy ultimately governs the set of responsible policy outcomes that arise in response to the data parity problem”).

A. Two Related Kinds of Conduct on Which Legal Claims Could be Brought: Refusal to Supply and Refusal to Deal

One type of case that enforcers could pursue under existing frameworks would involve alleging that platform companies' conduct with respect to their datasets amounts to an impermissible "refusal to supply."³⁰ In general, as the Federal Trade Commission notes, "a seller has the right to choose its business partners."³¹ In other words, firms can sell to whomever they like, provided that "the refusal is not the product of an anticompetitive agreement with other firms or part of a predatory or exclusionary strategy to acquire or maintain a monopoly."³² While it would admittedly be difficult to argue that a given technology firm's refusal to share its data with consumers, its content providers or its advertisers was part of such a strategy, there is U.S. Supreme Court precedent for treating data as an input to final products or services.³³ The enforcement agencies have also been willing to do so recently.³⁴ Thus, there is nothing that is so distinct about these markets that would prevent an enforcement agency or private plaintiff from bringing a refusal to supply case; the principal obstacles in such litigation are the same as they are in more conventional settings. Moreover, preventing their competitors from accessing or using their datasets could be interpreted as part of a firm's strategy to restrict competition, although these firms would likely raise either of the defenses discussed at the end of this section. Such a claim would also be able to gain traction only if the datasets involved are unique or in some way otherwise impossible to reproduce with sufficient speed by other firms further down the supply chain.

A second type of case that would similarly focus on a firm's conduct with regard to other firms is based on its "refusal to deal" with its horizontal competitors, when the dataset(s) involved is impossible or difficult to reproduce.³⁵ Just as is true in the "refusal to supply" context, firms have no general duty to deal with competitors. As enforcers acknowledge, doing so may even itself be an antitrust violation.³⁶ Antitrust liability, however, may exist if the firm that is refusing to deal with its competitors is itself a monopolist or has previously done business with its competitors and now stops.³⁷ Here again, however, the principal conceptual innovation that courts and enforcers must be willing to accept is to see a dataset as a product or service with the potential to be transacted from one firm to another.³⁸

B. Impacts on "Price" in the Market for Online Advertisements

A more novel antitrust litigation approach to improving access to the datasets on which AI development will depend might involve focusing on the portion of platform companies' business that engages with advertisers. The focus of such a strategy would be to allege that a lack of competition among large Internet platform companies harms consumers by increasing the quantity of advertising they must endure when accessing Internet content. This approach would admittedly take enforcers far afield from simply considering the competitive effects of lack of access to data. Nonetheless, the end result of disrupting platform company monopolies in this manner could still end up being a reduced foreclosure of dataset availability for competitors and entrants.

A simplified characterization of these markets is that firms purchase advertisements from platforms in much the same way as they do in conventional contexts (e.g. magazines or billboards). This characterization, however, is incomplete, owing to the two-sided or even multi-sided nature of these markets. The online advertising market instead occurs *across* the platform among consumers of the platform's content, the content providers and advertising firms. In such a conception, consumers receive desired content on the platform from the

³⁰ *Refusal to Supply*, Fed. Trade Comm'n, <https://www.ftc.gov/tips-advice/competition-guidance/guide-antitrust-laws/dealings-supply-chain/refusal-supply>.

³¹ *Id.*

³² *Id.*

³³ See *FTC v. Indiana Fed'n of Dentists*, 476 U.S. 447, 461 (1986) ("dentists'...refusing to supply the requested [patient X-ray] information [to dental insurers] was an unreasonable restraint of trade."). See also Stucke & Grunes, *supra* note 21, at 263; Patterson, *supra* note 27, at 15-16. This framework would get substantially more complicated, however, if data are viewed as intellectual property. See, e.g. U.S. Dep't of Justice & Fed. Trade Comm'n, Antitrust Guidelines for the Licensing of Intellectual Property, available at: <https://www.ftc.gov/sites/default/files/attachments/competition-policy-guidance/0558.pdf>.

³⁴ See, e.g. McSweeney, *supra* note 4, at 4 ("In the mergers involving big data that the FTC has investigated and challenged, the data is either a key input or the good or service itself").

³⁵ *Refusal to Deal*, Fed. Trade Comm'n, <https://www.ftc.gov/tips-advice/competition-guidance/guide-antitrust-laws/single-firm-conduct/refusal-deal>.

³⁶ *Id.*

³⁷ *Id.*

³⁸ Patterson, *supra* note 27, at 165 (noting Acting FTC Chairman Maureen Ohlhausen's past articulated support for treating data as both an input and an asset).

content provider, whom the advertiser is ultimately funding in exchange for consumers' time and attention to the advertising content as well as the relinquishing of their personal data to the platform company. While providing personal data has been described as the price of platform access in the past, consumer time and attention can also be thought of as part of the "price" paid for access to desired content.³⁹ Thus, when platform companies operate as monopolists, one might think that the amount of advertising endured by consumers would increase as well. In addition, quality of content and advertisements might diminish, as highlighted by the recent outcry over "fake news."⁴⁰

Presumably, consumers derive value from content access at a level that exceeds the amount of disutility they get from enduring advertisements, otherwise they would not visit websites hosting such content. This "revealed preference" idea, however, does not mean that these markets are achieving a first-best outcome without welfare losses such that no intervention is needed. Consumers routinely purchase goods and services from monopolists or other sellers operating in settings characterized by market failure, and yet there may still be a need for antitrust enforcement in such scenarios.⁴¹

Admittedly, firms would only continue to purchase advertisements on these platforms if a sufficient subset of consumers then makes purchases after having viewed these advertisements. For this subset of consumers, the benefits derived from the advertising market are obviously not limited to the content they can access as a result of having viewed the advertisements. The harm arising from a monopolized advertising market thus primarily falls on the consumers for whom the advertisement is simply to be endured or ignored. Although the benefits for these consumers of viewing desired content (e.g. search results or news stories) exceed their private costs of viewing advertisements, an online advertisement market that is lacking in platform company competition may fail to internalize fully these consumers' time costs of viewing advertising.

The costs of enduring more advertising of course do not take the form of direct financial harm — as in a conventional antitrust product market — but instead increase the amount of time consumers must spend accessing essential online services.⁴² In order for this kind of case to succeed, however, courts would have to be willing to accept the idea of treating attention or time as resources that can be measurable or at least construed as part of a theory of harm.⁴³ The difficulty in quantifying such effects may make courts more reluctant to embrace such a theory in spite of the academic literature's ability to do so.⁴⁴

C. Firms' Defenses: Privacy and Network Effects

Large platform companies might raise two defenses against any claim that the lack of competition in the market for large datasets is a violation of the antitrust laws. The first type of defense involves privacy. Firms could argue that they declined to share their data with competitors or potential downstream customers because of concerns that these entities would gain improper access to customers' data to which consumers never consented. While such a position appears to favor a greater degree of privacy protection for consumers, firms have no problem contracting around this limitation; firms often do so when it is in their interests by securing permission from consumers in a user agreement for their data to be shared.

39 See, e.g. Gentzkow, *Trading Dollars for Dollars: The Price of Attention Online and Offline*, 104 AM. ECON. REV.: PAPERS & PROC. 481 (2014); Teixeira, *The Rising Cost of Consumer Attention: Why You Should Care, and What You Can Do About It* 1 (Harvard Business School, Working Paper No. 14-055, 2014) http://www.hbs.edu/faculty/Publication%20Files/14-055_2ef21e7e-7529-4864-b0f0-c64e4169e17f.pdf ("the cost of attention has increased dramatically (seven- to nine-fold) in the last two decades").

40 Hubbard, *Fake News Is A Real Antitrust Problem*, ANTITRUST CHRONICLE (December 2017); Grunes, *Is "Fake News" A Competitive Problem?*, ANTITRUST CHRONICLE (December 2017).

41 This characterization may be especially apt if we acknowledge that access to some online services (e.g. e-mail, job search websites, online bill payment, etc.) is a non-negotiable feature of modern life, and viewing advertisements is required to gain access, unlike looking at billboards or reading a particular magazine. And if one were to argue that the market would simply provide websites with fewer or different advertisements if consumers wanted them, then that further reinforces the need to examine whether there is sufficient competition among major technology platforms, since they serve as the conduit that controls the terms by which both advertisers and content providers reach consumers. See also Stucke & Grunes, *supra* note 21, at 58-61 (highlighting additional problems with reliance on revealed preference arguments to justify the levels of quality in certain online markets).

42 These costs are bound to be borne both by consumers of lower sophistication — who may be unable to afford subscription, ad-free services or do not employ ad-blocking software — and by consumers of higher sophistication whose time costs are higher because of their intrinsically higher wages.

43 See generally Wu, *THE ATTENTION MERCHANTS, THE EPIC SCRAMBLE TO GET INSIDE OUR HEADS* (2016) (treating human attention as a commodity that firms are competing to attract).

44 See Teixeira, *supra* note 40, at 1.

A second argument that platform companies might make is that monopolies over these large datasets give rise to efficiencies for both other firms and consumers by virtue of network effects. These positive externalities make it so that the value of the firm's data increases with the size of the dataset; it is not only the individual observations that generate value but also the broader context in which they are situated.⁴⁵ Consumers benefit from having access to a wide range of data about themselves as well as other consumers. Other firms — from advertisers to content providers — benefit from being able to know as much as possible about a single individual. While this argument explains why it is beneficial for one firm to have large portions of the data that are available about an individual — rather than have it be allocated across firms — network externalities alone fail to fully justify why a single firm should be able to refrain from making data interoperable or portable among firms. Firms may make the additional argument here that this monopoly is their hard-earned reward for creating and maintaining a platform that benefits such a large volume of individuals. But in a market where the relationship between competition and innovation more closely matches Arrow's account than Schumpeter's, this argument also does not provide a justification for preventing other firms from having access to this data; enabling competition would not be expected to reduce the incentives for innovation.

V. POTENTIAL POLICY AND REGULATORY SOLUTIONS

In this section, we describe a variety of proposals that aim to ultimately increase the amount of innovation being done by AI firms. The ideas vary as to whether they help to increase competition *for* the market by encouraging new entrants into the market, or whether they help to increase competition *in* the market by enabling customers to more easily switch between existing firms.

A. *Deferred Data-Sharing Requirement: Temporary Data Monopolies Followed by Complete Data Availability*

One potential regulatory solution to the lack of competition among platform companies with large datasets would be a data-sharing requirement for datasets above a certain size that would presumptively go into effect after a fixed amount of time from the date on which the data were first collected. Prior to that date, firms would be free to restrict access to their data, but any unlawful dealings with other firms — either by refusing to share datasets or sharing them inappropriately — would still be litigable under the antitrust laws. Such a solution is analogous to the data exclusivity provisions for biologic pharmaceuticals, which under current law last for 12 years as specified under the Biologics Price Competition and Innovation (“BPCI”) Act of 2009.⁴⁶

The idea behind such a solution is that by maintaining an individual firm's ability to secure a temporary monopoly, it would preserve a key incentive for investment. At the same time, the eventual sharing requirement would — just like finite patent and copyright terms⁴⁷ — provide other AI innovators with the ability to freely access and use the raw material of innovation in the long run. This setup would be particularly useful for prospective or recent entrants into the marketplace, as they would be less likely to be able to benefit from the blockchain-based solution described below, in contrast with their incumbent firm counterparts.

The difficult part in designing and administering such a measure involves setting the right time horizon. If the amount of time until the sharing requirement is activated is too short, then the incentives for investment are dampened. There may also be privacy concerns if certain datasets are made public too quickly. On the other hand, if the time horizon is too long, these data may no longer even be useful to innovators.⁴⁸ Setting the time horizon will likely be the result of a stakeholder engagement process that weighs the competing concerns of many constituencies, as was the case with the BPCI Act.

In addition to setting an appropriate time horizon, it would probably also be necessary to determine a carve-out or waiver process by which data-owning firms can demonstrate that their holdings should be exempt from an automatic disclosure requirement, especially on either privacy or trade secret grounds. Another way of handling such sensitive datasets would be to design the requirement such that there is an intermediate level of disclosure available, e.g. disclosure to a third-party repository that the general public is unable to access but that is open to innovators who apply and are vetted (the use of third parties is described in detail in the next section).

45 Stucke & Grunes, *supra* note 21, at 23 (“[I]n some industries, simple algorithms with lots of data will eventually outperform sophisticated algorithms with little data.”).

46 Biologics Price Competition and Innovation Act of 2009, 42 U.S.C. § 262(k)(7)(A).

47 This analogy is imperfect, however. In both the patent and copyright contexts, other innovators can usually at least see and examine the protected work from the beginning. By contrast, in the market for data, simply viewing a dataset would almost always enable its use, so in order for this solution to be practicable, both viewing and using can only take place at the same time after the sharing requirement kicks in.

48 Stucke & Grunes, *supra* note 21, at 21 (“[T]he older the data, the less valuable it is.”).

B. Data Portability and the Use of Trusted Third Parties

Another possible solution is the use of a data portability mechanism.⁴⁹ Under such a model, a customer would maintain possession of some core data about herself that she could then take from one company to a rival, in much the same way that a phone customer can take her phone number from one provider to another. In principle, this measure should help increase competition between established firms *in* the market, because any potential customer could easily shift her data from one established firm to another. However, it is unlikely that data portability alone would increase competition *for* the market. Startups that want to enter a market need access to large datasets to train their AI algorithms, and it seems impractical to expect a startup to assemble such a dataset by relying on individual users porting their data to the startup in a piecemeal fashion.

Another issue with data portability is where the customer's data would reside, which has implications for the data's security. One possibility is for the data to reside with a trusted third party, such as an educational institution, or perhaps an organization created via a public-private partnership. A key role for a third party is to protect the privacy and security of the data, while allowing for other parties to access it conditional on approval. An appealing feature of a trusted third party is that once the data are anonymized, they could potentially be combined with other data for use by entrants to train their AI, with the result that these datasets are not assembled piecemeal. A common technique currently used by entrants to overcome the lack of customer data is to train their AI on publicly available datasets. But if these datasets are biased in some way, then the resulting AI algorithms will reflect the bias. The worry is that if many entrants use similarly biased datasets, then bias quickly propagates.⁵⁰ Tom Mitchell and Erik Brynjolfsson argue for the collection and integration of AI-related data from diverse sources, a trusted broker to summarize and protect the privacy and security of the data, and normalization of the data where possible to address any different skews and biases across different datasets.⁵¹ They hope that, combined with existing measures, this information infrastructure can provide a comprehensive picture of the true effects of technological advancement, thus allowing decision-makers to respond effectively.

There are several good examples of educational institutions working together to house data for use by other researchers, including ICPSR, IRIS and NORC. Educational repositories of data serve multiple purposes, notably the ability of researchers to replicate each other's studies, or to combine data in creative ways to answer new, complex questions.⁵² Traditionally, educational consortia have not been used to house private party data, but the infrastructure and expertise is there. In short, the use of trusted third parties lowers the costs for entrants to access good training data, which should increase the quantity and quality of entrants, thereby leading to more innovation. However, this type of solution does not provide any control over data to customers.

C. Data Portability and Blockchain

Blockchain has been described as a distributed public ledger — a chronological list of transactions that is verified at regular intervals by shared users.⁵³ Christian Catalini and Joshua Gans, among others, have suggested that blockchain may have promising applications when it comes to consumer data. For example, Catalini and Gans write that:

[f]rom a privacy perspective, the ability to license out subsets of personal information for limited amounts of time and to seamlessly revoke access when necessary has the potential to not only increase security, but also to enable new business models where customers retain greater control over their data and firms can dynamically bid for access.⁵⁴

Thus, in the future, blockchain could lower the costs for customers to control and trade their data, which should increase competition between incumbent firms *in* the market, leading to benefits for consumers and potentially more innovation. However, while such a solution

49 Several officials and observers have suggested versions of this approach recently. See, e.g. Pegoraro, *Web Companies Should Make It Easier to Make Your Data Portable: FTC's McSweeney*, USATODAY, <https://www.usatoday.com/story/tech/columnist/2017/11/12/web-companies-should-make-easier-make-your-data-portable-ftcs-mcsweeney/856814001/> (highlighting FTC Commissioner Terrell McSweeney's calls for such measures).

50 Levendowski, *How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem*, WASH. L. REV. (forthcoming).

51 Mitchell & Brynjolfsson, *Track How Technology is Changing Work*, 544 NATURE 290-91 (2017).

52 Lane, *Uses of Micro-data*, Keynote Speech at the Conference of European Statisticians (June 12, 2003), <https://www.unece.org/fileadmin/DAM/stats/documents/ces/2003/crp.2.e.pdf>.

53 HALABURDA & SARVARY, *BEYOND BITCOIN: THE ECONOMICS OF DIGITAL CURRENCIES* (2015); Watkins & Rodriguez, *WTF is Blockchain?!*, OZY, <http://www.ozy.com/fast-forward/wtf-is-blockchain-inside-the-most-disruptive-tech-since-the-internet/81567>.

54 Catalini & Gans, *Some Simple Economics of the Blockchain* 22 (Nat'l Bureau of Econ. Research, Working Paper No. 22952, 2016).

likely improves upon security and portability of data relative to a trusted third party, it may not be helpful at increasing entry by startups. As with other data-portability solutions, unless the startup is able to assemble a really large dataset that it can use to train its AI, these solutions are of limited value when it comes to increasing competition *for* the market.

VI. CONCLUSION

Artificial intelligence is already having a dramatic impact on our economy and society. Like any other technology, competition between AI firms can lead to many new and useful innovations that benefit consumers, but the need for datasets to enable AI firms may be a barrier to entry. If so, and if current antitrust approaches are not able to address these barriers for the ultimate benefit of future AI consumers, then creative policy and regulatory solutions are needed. We have attempted to lay out a few of these solutions above, while discussing some of the tradeoffs inherent in each of them.

