Free markets. Real solutions.

R STREET POLICY STUDY NO. 153
October 2018

REDUCING ENTRY BARRIERS IN THE DEVELOPMENT AND APPLICATION OF AI

Caleb Watney

CONTENTS

INTRODUCTION

Artificial Intelligence (AI) is developing rapidly and countries from around the globe are beginning to articulate national strategies for handling the political ramifications.[1] Powering innovations like driverless cars, autonomous drones, full-sequence genetic analytics and powerful voice assistant technology, the future certainly looks full of potential.[2] However unsettled questions about who will reap these benefits and when they will be achieved leave storm clouds on the political horizon. Amid questions of industrial concentration and economic inequality on one side,[3] and concerns about lagging U.S. productivity and the slow pace of AI diffusion on the other,[4] this paper attempts to lay out a framework that can begin to address these various issues.

The first of these questions could be simplified to ask: what if only Google, Amazon, Facebook and Apple (GAFA) are able to develop the AI system that powers the economy of the future? The second considers the reasons that AI techniques have diffused through the economy at such a slow rate. However, although these appear to be two distinct questions, there is an under-examined overlap that connects these issues to the same set of policies: namely, high barriers to entry due to supply- side constraints.

There are significant barriers to entry in AI development and application, many of which stem from the direct result of government policies. These barriers have inadvertently boosted the market power of incumbent firms and thus in reducing them, we may enable new firms to better compete, while also removing some of the bottlenecks that slow down research and integration of AI systems across the entire economy.

Accordingly, the present study provides an overview of the various inputs to the production function of AI and analyzes the policies that should be reconsidered or implemented to reduce these barriers. Specifically, it will focus on the inputs of skilled AI analysts, high-quality datasets and specialized AI hardware. It will conclude with a short discussion of the relative attractiveness of focusing on entry barriers when

1. See, e.g., Tim Dutton, "An Overview of National AI Strategies," *Politics + AI*, June 28, 2018. https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd.

2. For the exciting potential of AI in speeding the rate of economic growth and innovation, see, e.g., Iain Cockburn et al., "The Impact of Artificial Intelligence on Innovation," *National Bureau of Economic Research Working Paper* No. 24449, March 2018. http://www.nber.org/papers/w24449.

3. See, e.g., Kai-Fu Lee, "The Real Threat of Artificial Intelligence," *The New York Times*, June 24, 2017. https://www.nytimes.com/2017/06/24/opinion/sunday/artificial-intelligence-economic-inequality.html.

4. Erik Brynjolfsson, "Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics," *National Bureau of Economic Research Working Paper* No. 24001, November 2017. http://www.nber.org/papers/w24001.

compared to the high-risk options presented by traditional antitrust enforcement. While there are certainly other potential policies or AI inputs that are beyond the scope of this paper, the policy framework presented herein will nevertheless provide a useful primer for future analysis.

A NOTE ON TERMS

At the outset, it is helpful to define a few specific terms that are applied in the following analysis. "AI" is meant to refer broadly to the set of computer algorithms being used to automate or improve aspects of human decision-making.[5] In the most current iteration, this is largely being accomplished via machine learning (ML), whereby an algorithm uses statistical techniques to progressively improve prediction ability for a given task.[6] By this definition, AI exists on a spectrum rather than as a binary, with increasing sophistication in the ability to apply various models to solve the problem at hand indicating higher levels of intelligence.

The term "AI development" refers to the research process of creating more advanced algorithms on the technological frontier. By contrast, "AI application" denotes the implementation of AI systems that have already been developed to new industries and problems. Development is vital for advancement in the field, while application is necessary for those advancements to actually affect the economy.

SUPPLY OF SKILLED AI ANALYSTS

Perhaps the single biggest bottleneck in AI development and application today is the supply of skilled data scientists and machine-learning engineers. Typical AI specialists can expect to earn between $300,000 and $500,000 at top tech firms; numbers that are significantly higher than their peers in other computer-science-related subfields.[7] In addition to these ballooning salaries, industry experts like Hal Varian have pointed to the scarcity of adequate AI talent as the largest factor behind slow application in the economy.[8]

While the number of individuals pursuing careers as skilled AI analysts has certainly been increasing, the length of time it takes to develop necessary technical skills and the surging demand for AI specialists have created an intense labor shortage that benefits large, established firms. When bidding against deep-pocketed incumbents who can afford to pay the high six-figure salaries required to be competitive, it is difficult for startups and smaller businesses to compete for limited talent.[9] Similarly, given the costs of acquiring a skilled team for AI application, even established firms in non-tech sectors that may be able to afford high compensation, will face a high bar to experimentation. So long as AI talent is sufficiently limited, it seems likely that the existing supply will be funneled primarily toward development rather than application.

For this reason, if there were appropriate policy levers to increase the supply of skilled technical workers available in the United States, it would disproportionately benefit smaller companies and startups. This would make the overall ecosystem more competitive while simultaneously increasing the rate of AI diffusion in other industries. To accomplish this, the following proposals should be considered.

Reform our immigration system to allow more high-skill AI talent

The policy lever with perhaps the highest degree of leverage to begin immediately alleviating this talent shortage is our immigration system and more specifically, reform around visas for international graduate students.

In 2015, the United States had 58,000 graduate students in computer science fields, the overwhelming majority of which (79%) were international.[10] This represents a significant portion of the overall AI talent supply being cultivated each year, as students from all over the world are attracted to the nation's top education system. In particular, the United States attracts large numbers of students from China and India.[11] However, due to a limited number of visa slots, only a fraction of these students is allowed to work in the country long term.[12]

The primary pathway for these highly skilled immigrants to stay in the country is through the H-1B visa program.[13] However, for the past 16 years, the H-1B limit has been exhausted and, in more recent years, the number of applications filed has consistently been twice as high as the number of avail-

5. While definitions of AI vary, for an overview, see Peter Stone et al., "Artificial Intelligence and Life in 2030: One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel," *Stanford University*, September 2016. p. 12. https://ai100.stanford.edu/2016-report.

6. Ibid, pp 2-4.

7. Cade Metz, "Tech Giants Are Paying Huge Salaries for Scarce A.I. Talent," *The New York Times*, Oct. 22, 2017. https://www.nytimes.com/2017/10/22/technology/artificial-intelligence-experts-salaries.html.

8. Hal Varian, *The Economics of Artificial Intelligence: An Agenda* (University of Chicago Press, Forthcoming), p. 20. http://www.nber.org/chapters/c14017.pdf.

9. See e.g., Michelle Cheng, "How Startups Are Grappling With the Artificial Intelligence Talent Hiring Frenzy," *Inc*., May 25, 2018. https://www.inc.com/michelle-cheng/how-startups-are-grappling-with-artificial-intelligence-talent-hiring-frenzy.html.

10. "The Importance of International Students to American Science and Engineering," National Foundation for American Policy, October 2017. http://nfap.com/wp-content/uploads/2017/10/The-Importance-of-International-Students.NFAP-Policy-Brief.October-20171.pdf.

11. Ibid, p. 14.

12. "H-1B Visas by the Numbers: 2017-2018," National Foundation for American Policy, April 2018. https://nfap.com/wp-content/uploads/2018/04/H-1B-Visas-By-The-Number-FY-2017.NFAP-Policy-Brief.April-2018.pdf.

13. Ibid.

able spots.[14] And this is almost certainly understating the scope of the problem, as it does not account for the ways in which foreknowledge about the difficulty of acquiring a work visa may deter students from applying in the first place.

Although it also limits the talent pool available to large tech firms, the status quo is especially daunting for startups, as they do not have the specialized Human Resources personnel to handle the bureaucracy of the immigration visa application process. Including application and attorney fees, to sponsor a work visa typically costs around $5,000 per employee[15] and the paperwork burdens appear to be increasing.[16] Both the financial and bureaucratic costs are easier for established firms to bear, given their larger size and increased resources.[17]

In turn, this impacts the types of firms high-skill immigrants will apply to work for in the first place. Even when attracted to work at startups, foreign workers may ultimately privilege their applications to incumbents because they will likely have a better chance of obtaining work visas at established firms. Additionally, since startups face high failure rates, job loss could mean termination of work authorization as well. This would mean that the entire visa application process would have to be approached anew. Indeed, a recent longitudinal survey concluded:

> *Although foreign [STEM PhD students] are 45% more likely to be interested in working in a startup prior to graduation [when compared to US students], after graduation they are 50% less likely to do so. Controlling for ability and other characteristics, ex ante career interests are a strong predictor of startup employment among U.S. workers but not among foreign workers, suggesting that foreign workers may face constraints in choosing their preferred jobs [...] suggesting a potential pool of entrepreneurial labor that might move to startups if provided the opportunity to do so.*[18]

Accordingly, to allow more international students to live and work in the United States upon completion of their degree—either through an expansion and simplification of the H-1B visa program or through the creation of a new technical worker visa program—would be a relatively straightforward and effective method to alleviate the country's talent shortage around AI. In particular, this would benefit smaller firms and startups that are unable to access existing foreign-born talent to the same degree as established firms.

Allow companies to deduct the cost of training AI talent

In addition to reforming our immigration pathways for high-skilled AI talent, it would be wise for the United States to extend more effort toward building up domestic talent. However, given that it can take years to train new AI specialists when compared with the near-instant effect of allowing already-trained, foreign-born experts to stay in the country, this will likely require a longer timeframe for the resources spent on this approach to pay off.

As the number of newly minted machine learning PhD students continues to dwindle, some companies are looking at training employees internally to essentially create new supply.[19] However, it requires significant investment on the company's part, both in time and resources, to train new AI specialists this way, and the gains from this training are mostly captured by the newly trained worker in the form of higher wages. In light of this and since workers can jump ship from the companies that train them at any time for a higher salary at a competitor, employers have few opportunities to recoup the costs of worker training.[20] It thus seems likely that employers are generally underinvesting in worker training when compared to the amount that might otherwise be efficient. We should therefore look more closely at incentivizing this socially desirable behavior through the tax code.

Employers may currently deduct a portion of the costs of worker training as long as it is to improve productivity in a role they already occupy, but this credit is fairly small and employers may not deduct the costs if it would qualify them for a new trade or business.[21] Expanding this deduction—both in size and scope—so that the full cost of worker training for new trades could be deducted, would incentivize more investment in building the AI workforce that is needed

14. Ibid, p. 2.

15. Matt Faustman, "How Much Will Sponsoring an H1-B Visa Cost an Employer?", *Upcounsel Blog*, 2013. https://www.upcounsel.com/blog/what-is-the-costs-for-an-employer-to-sponsor-an-h1b-visa.

16. Ana Campoy, "Trump is quietly swamping visa applicants in extra paperwork," *Quartz*, Jan. 11, 2018. https://qz.com/1176576/h1b-visa-under-trump-is-already-harder-to-get.

17. Jana Kasperkevic, "Getting an H-1B visa is becoming more difficult," *Marketplace*, April 2, 2018. https://www.marketplace.org/2018/03/30/business/immigration-reform-affect-businesses-hiring-visa-workers.

18. Cited with permission from Michael Roach et al., DRAFT: "U.S. Immigration Policies and the STEM Entrepreneurial Workforce," National Bureau of Economic Research, April 2018. http://www.nber.org/chapters/c14101.pdf.

19. "So we invite folks from around Google to come and spend six months embedded with the machine learning team, sitting right next to a mentor, working on machine learning for six months, doing some project, getting it launched and learning a lot." See, Steven Levy, "How Google is Remaking itself as a 'Machine Learning First' Company," *Wired*, June 22, 2016. https://www.wired.com/2016/06/how-google-is-remaking-itself-as-a-machine-learning-first-company.

20. See e.g., Alastair Fitzpayne and Ethan Pollack, "Worker Training Tax Credit: Promoting Employer Investments in the Workforce," The Aspen Institute, May 12, 2017. https://www.aspeninstitute.org/publications/worker-training-tax-credit-promoting-greater-employer-investments-in-the-workforce.

21. Michael Farren, "Bridging the Skills Gap," Congressional Testimony before the House Small Business Committee, Subcommittee on Economic Growth, Tax, and Capital Access, Examining the Small Business Labor Market, Sept. 7, 2017. https://www.mercatus.org/system/files/farren_-_testimony_-_bridging_the_skills_gap_-_v2.pdf.

to fuel our economy.[22] Given the pre-existing level of interest by employers in this strategy, it seems likely this could become a fruitful part of our domestic AI pipeline, if given more support.

Another way to frame this issue is by comparing the incentives for investment in worker training with those in other areas. For example, unlike investments in human capital development, investment expenditures for capital goods—like factories or robots—are currently fully deductible in the tax code.[23] This creates a system that incentivizes employers to invest more in capital productivity gains rather than labor productivity gains, which should be equalized to create a fairer playing field.[24]

As a simple example of a way in which this change could increase the supply of AI talent and speed AI diffusion, consider a hypothetical owner of a manufacturing plant. This owner has information technology (IT) staff who are generally technologically competent, but possess no special training in machine learning. She might be interested in sending this staff to a six-month, ML boot camp where they could learn the basics of applying ML techniques to analyze production processes and find new efficiencies in her manufacturing plant.[25] However, currently, such an expense would not be deductible, potentially discouraging her from making such an investment in the first place.

All else equal, allowing the costs of worker training to be fully deductible will spur more worker training. In the case of companies both developing AI and companies that could benefit from AI application, this means increasing the overall supply of skilled AI analysts in the economy. These workers will likely go on to use their skillsets for future employers, helping spur productivity growth and making the overall ecosystem more competitive.

22. In economic terms, an externality refers to a side effect or consequence of private sector action for which the effects are not fully reflected in the cost of the good or service. In this case, the positive effects of increased AI-talent supply for the entire economy are not fully internalized by the companies individually training workers. For this reason, many may generally be undersupplied.

23. See e.g., Gabriel Horwitz, "How The Government Perversely Encourages Machine Over Human Capital," *Forbes*, March 28, 2017. https://www.forbes.com/sites/washingtonbytes/2017/03/28/how-the-government-perversely-encourages-physical-over-human-capital/#23016623f9c6.

24. Some commenters have argued that to fully equalize the playing field between capital and labor improvement would require full-fledged, human-capital tax credits, similar to research and development tax credits. See, e.g., Rui Costa et al., "Investing in People: The Case for Human Capital Tax Credits," Centre for Economic Performance, Paper ISO1, February 2018. http://cep.lse.ac.uk/pubs/download/is01.pdf.

25. As an example of the type of programs more employers might take advantage of if the expense was tax deductible, see e.g., Austin Allred, *Twitter*, Sept. 12, 2018. https://twitter.com/AustenAllred/status/1039921578904043520

SUPPLY OF DATA

In many ways, the supply of high-quality machine-readable training data is the key enabler of machine learning. Without access to some underexplored dataset, a team of talented AI specialists can be left twiddling their thumbs. Consumer data in the United States is particularly valuable but large firms like GAFA have underlying digital services that supply them with immense reams of valuable and unique consumer data. Competitors do not easily have the same access.[26]

This is not inherently an issue, as these large technology companies have invested billions of dollars to create services that provide significant value for consumers and in return, consumers have shown a willingness to contribute their data.[27] We should aspire for other companies to create services that prove to create as much value for consumers. However, it is undoubtedly an advantage in particular domains of AI work that startups are currently unable to replicate.[28]

We should be careful to note, however, that beyond a certain threshold, increases in the sheer volume of data possessed generate decreasing returns to scale.[29] This means that while possessing high-quality data is vital for performance, simply having more data than a competitor is no guarantee of victory.[30] In fact, we are seeing that the role of sophisticated algorithmic design and ML feedback loops is only increasing.[31] Sometimes a smaller competitor with an adequate dataset and insightful algorithmic design can outperform an incumbent with a superior dataset but mediocre design.

Given all this, we can potentially create high-leverage opportunities for startups to compete against established firms if we can increase the supply of high-quality datasets available to the public. As with increasing the supply of AI talent, this will help both incumbents and startups but on the margin, it will be the smaller firms with less access to consumer data who benefit most.

26. See e.g., Doug Aley, "It's Hard to Compete With Tech Giants Like Google and Amazon—But It Can Be Done," *Entrepreneur*, July 18, 2018. https://www.entrepreneur.com/article/316376.

27. Erik Brynjolfsson et al., "Using Massive Online Choice Experiments to Measure Changes in Well-being," *National Bureau of Economic Research Working Paper* No. 24514, April 2018. http://www.nber.org/papers/w24514.

28. See, e.g., Tom Simonite, "AI and 'Enormous Data' Could Make Tech Giants Harder to Topple," *Wired*, July 13, 2017. https://www.wired.com/story/ai-and-enormous-data-could-make-tech-giants-harder-to-topple.

29. For each piece of data accumulated, the amount of predictive power acquired decreases. See, e.g., "Stanford Dogs Dataset," Stanford University, 2011. http://vision.stanford.edu/aditya86/ImageNetDogs.

30. See, e.g., Joe Kennedy, "The Myth of Data Monopoly: Why Antitrust Concerns About Data Are Overblown," Information Technology and Innovation Foundation, March 2017. http://www2.itif.org/2017-data-competition.pdf.

31. See, e.g., Xavier Amatriain, "In Machine Learning, What is Better: More Data or better Algorithms," *KDnuggets*, June 2015. https://www.kdnuggets.com/2015/06/machine-learning-more-data-better-algorithms.html.

Encourage the creation of open datasets and data sharing

One of the easiest ways to begin this process would be a more thorough examination of existing government datasets that are not public. As an example of previous projects that were broadly successful, consider the U.S. National Oceanic and Atmospheric Administration (NOAA) and Landsat projects, both of which made weather-satellite data available to the public and, in turn, developed into a multi-billion-dollar industry, creating more accurate forecasts of extreme weather and crop patterns.[32]

There appears to be even more potential from datasets the government owns but has not made public. For example, many cities and municipalities have useful data around traffic patterns, electricity usage and business development that, if made accessible, could lead to reduced-cost service provision and better analytics.[33] And there have been a flurry of recent pushes in Congress to standardize the publication of government agency datasets in a machine-readable format.[34]

It is frequently difficult to know beforehand how new data will be leveraged by startups and what new industries might form around it. After all, when the U.S. Government first made GPS-satellite data available to the public, they had no idea it would eventually become the backbone for location-tracking services used in smartphones around the world.[35] This should lead to a general presumption in favor of releasing government data, even if the consumer applications do not appear immediately obvious.

While there has been some concern around the privacy implications of making more government data public, especially data that might become personally identifiable, recent advances in data anonymization techniques like differential privacy should lessen these concerns.[36] While there may still be data that would be inappropriate to release to the public for national security or privacy reasons, it appears there is still significant progress to be made at current margins.[37]

There is also the matter of industries in which open data might become the norm if existing regulations are relaxed or streamlined. The healthcare industry seems a particularly promising target in this respect, as HIPAA has long been considered a barrier to the development of data sharing between medical professionals and companies.[38] Allowing consumer health data to be more easily shared with the proper privacy safeguards could enable a renaissance in drug development and personalized medicine, as recent ML advances have proven quite promising when appropriate data have been available.[39]

Each new dataset that can be easily shared or, when appropriate, made public, increases the odds both that a new startup will be able to leverage it for success, and also that a new industry can thrive around the increased predictive analysis the released data has enabled. For recent advances in AI to diffuse throughout the economy, we must make sure the underlying data is accessible.[40]

Clarify the fair-use exemption for training data

In addition to making more government datasets open source, we should also take a second look at some of the intellectual property laws that intersect and interact with the ML process, specifically copyright law.

Imagine a hypothetical startup focused on the creation of a natural-language-processing application. One readily available source of human dialogue the company might consider learning from would be the last 50 years of Hollywood scripts, many of which are scrapable from various online databases. However, such an endeavor would stand on legally dubious grounds, as these scripts remain copyrighted works and there have not been clear legal guidelines established to delineate what is allowable as fair use in ML training data. Given this, it is more likely that such a startup would avoid

32. See, e.g., Christina Rogawski et al., "NOAA Open Data Portal: Creating a New Industry Through Access to Weather Data," *Open Data's Impact*, January 2016. http://odimpact.org/files/case-studies-noaa.pdf; and Tom Lee, "Closing Landsat data is (still) a bad idea," *Medium*, Aug. 9, 2018. https://medium.com/@thomas.j.lee/closing-landsat-data-is-still-a-bad-idea-8ef0ccfcc7dc.

33. See, e.g., Michael Chui et al., "Innovation in Local Government: Open Data and Information Technology," McKinsey Global Institute, 2014. https://goo.gl/wfSsro.

34. See, e.g., S.2852, "OPEN Government Data Act," 114th Congress, April 26, 2016. https://www.congress.gov/bill/114th-congress/senate-bill/2852.

35. Andrew Young et al., "United States GPS System: Creating a Global Public Utility," *Open Data's Impact*, January 2016. http://odimpact.org/files/case-studies-gps.pdf.

36. See, e.g., Kobbi Nissim et al., "Differential Privacy: A Primer for a Non-technical Audience (Preliminary Version)," *Vanderbilt Journal of Entertainment and Technology Law*, April 13, 2018. https://privacytools.seas.harvard.edu/publications/differential-privacy-primer-non-technical-audience-preliminary-version.

37. For more case studies of successful open-data initiatives and lessons learned, see Stefaan Verhulst and Andrew Young, "When Demand and Supply Meet: Key Findings of the Open Data Impact Case Studies," *Open Data's Impact*, March 2016. http://odimpact.org/files/open-data-impact-key-findings.pdf.

38. Niam Yaraghi, "To Foster Information Exchange, Revise HIPAA and HITECH," *Health Affairs*, Sept. 19, 2017. https://www.healthaffairs.org/do/10.1377/hblog20170919.062032/full.

39. See, e.g., Rob Matheson, "Artificial intelligence model 'learns' from patient data to make cancer treatment less toxic," *MIT News Office*, Aug. 9, 2018. http://news.mit.edu/2018/artificial-intelligence-model-learns-patient-data-cancer-treatment-less-toxic-0810; and Dave Gershgorn, "If AI is going to be the world's doctor, it needs better textbooks," *Quartz*, Sept. 6, 2018. https://qz.com/1367177/if-ai-is-going-to-be-the-worlds-doctor-it-needs-better-textbooks.

40. Note that this would also imply that new overly restrictive privacy laws could have the effect of raising barriers to entry and slowing innovation. Indeed, empirical evidence to date would appear to confirm this. See, e.g., Ajay Agrawal et al., "Economic Policy for Artificial Intelligence," *National Bureau of Economic Research Working Paper* No. 24690, June 2018, pp. 9-10. http://www.nber.org/papers/w24690.

this potential legal minefield and consider what other datasets might be available with less risk.

This is the ambiguous state of copyright enforcement in ML today. Legal scholar Amanda Levendowski has argued that this de facto privileging of frequently low-quality data that exists in the public domain (like the Enron emails) has inadvertently biased the many AI applications that are built upon them.[41]

However, this may also have important and underexplored applications for the state of competition. There are an enormous number of copyrighted works that are scrapable from the Internet, the data of which is currently underexploited in part because of its legally dubious standing if used as training data. This could represent, then, a significant lever to create new arbitrage opportunities for scrappy startups willing to find and leverage interesting datasets. The full scope of what this might entail or lead to is admittedly difficult to fully grasp, considering the massive amount of data that might be included.

Google has already showcased one use case for which this type of data might be leveraged. In 2016, a research division within Google used a collection of 11,000 free e-books to show the potential improvements that could be made to a conversational AI program.[42] This sparked considerable controversy with groups like the Authors Guild who considered it a violation of the author's intended purpose and argued it was a copyright violation.[43] Because this was a research paper and not publicly used for later commercial purposes, no suit was pursued. Notably, however, the original "BookCorpus" dataset is no longer publicly hosted.[44]

Given the existing ambiguity around the issue and the large potential benefits to be reaped, further study and clarification of the legal status of training data in copyright law should be a top priority when considering new ways to boost the prospects of competition and innovation in the AI space.[45]

41. Amanda Levendowski, "How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem," *Washington Law Review* 93 (July 19, 2018), pp. 579-631. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3024938.

42. Samuel R. Bowman et al., "Generating Sentences from a Continuous Space," Google Brain, May 12, 2016. https://arxiv.org/pdf/1511.06349v4.pdf.

43. See, e.g., Richard Lea, "Google swallows 11,000 novels to improve AI's conversation," *The Guardian*, Sept. 28, 2016. https://www.theguardian.com/books/2016/sep/28/google-swallows-11000-novels-to-improve-ais-conversation.

44. See, for example, this GitHub forum discussion about the missing 'BookCorpus' dataset and the encouragement to scrape the data again oneself. https://github.com/ryankiros/neural-storyteller/issues/17.

45. For a more critical examination of the potential problems with expanding the scope of fair use in machine-learning training data, see Benjamin Sobel, "Artificial Intelligence's Fair Use Crisis," *Columbia Journal of Law & the Arts*, Forthcoming. https://ssrn.com/abstract=3032076.

ACCESS TO SPECIALIZED HARDWARE

Underlying the data being used to train ML models and the data scientists who are building them is the physical infrastructure of the AI world. This primarily takes the form of the computer servers and chipsets that ML models are trained and operated on. In recent years, this hardware has become increasingly specialized to keep up with the pace of AI development. As the tasks asked of various ML systems continue to diverge, the type of computational power enabled by specific chip architectures has become just as important as the sheer magnitude.

As AI scholar Tim Hwang has noted, there are two dynamics that are shaping the marketplace for ML hardware today.[46] The first is the inverse relationship between performance and flexibility or in other words, that general purpose hardware that tends to be less expensive and is used for a wide variety of computing tasks is being outpaced in performance by chipsets built for a specific task.[47] The second dynamic pertains to the differing types of hardware used for initial training of an ML model and for making inferences with an already-trained model.[48] For example, energy consumption for a computer-vision system may matter a great deal when operating on a mobile phone, but not when the model is originally being trained in a data center. As Hwang concludes: "These considerations influence what kinds of hardware are used at which points in the lifecycle of a machine learning system. They can be viewed as separate though overlapping markets, with hardware platforms being offered either for training or inference, and some offering support for both."[49]

While a natural and necessary part of the AI development process, such a trend toward specialized hardware does increase the fixed costs required to be competitive. This manifests not only in the expense of these systems, but in the elaborate supply chains that have been built up to support them. While the policy recommendations that flow out of this insight are less clear cut than those for the supply of AI analysts or datasets, maintaining access to valuable AI hardware is a key policy consideration.

Avoid causing political instability to international supply chains

As AI hardware becomes more specialized, the supply chains for very specific chips become a critical ingredient for cutting-edge ML research. While the United States maintains advanced manufacturing facilities for particular parts of the

46. Tim Hwang, "Computational Power and the Social Impact of Artificial Intelligence," *MIT Media Lab*, March 23, 2018. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3147971.

47. Ibid., p. 8.

48. Ibid., p. 9.

49. Ibid.

production process that are vital to the supply chain (like semiconductor fabrication), much of the production has been outsourced. Given the importance of chip foundries in Taiwan and China in particular, the perceived stability of trade in the region will alter investment patterns and domestic access to these sophisticated chips.[50]

To ensure access in spite of political tensions, large companies like Apple, Google and Nvidia are beginning re-shore production of especially valuable chips.[51] However, smaller competitors and startups are much more limited in this capacity and thus are more reliant on existing international supply chains.

Insofar as recent U.S. trade tensions with China have increased the perceived instability of regional trade, the disparate impact this will have on smaller firms should be recognized.[52] Ultimately, new foundries and semiconductor manufacturing plants will shift wherever it is most profitable. Accordingly, in the event of a long-term trade war, production could eventually shift elsewhere. However, it will certainly shape short- and medium-term access to specialized hardware.

While this analysis has focused on the effects to domestic competition, the pros and cons of a coordinated national security push to on-shore semiconductor manufacturing are beyond the scope of this paper, but the effects of that decision could impact the degree to which this continues to be a meaningful issue.[53]

Maintain a healthy ecosystem around distributed platforms

The other significant trend in AI hardware utilization is the growth of cloud-computing platforms like Amazon Web Services (AWS) and the Google Cloud platform. Cloud computing has notable pro-competitive effects in that it transforms what is normally a fixed cost in server capacity into a variable one.[54] Allowing a startup to buy only the discrete server space they will need for that month significantly reduces the amount of venture capital needed to get a company off the ground.

50. Ibid., pp. 18-21.

51. See, e.g., Reinhardt Krause, "In AI Technology Race, U.S. Chips May Be Ace-In-The-Hole Vs. China," *Investor's Business Daily*, Nov. 27, 2017. https://www.investors.com/news/technology/ai-technology-u-s-chip-stocks-vs-china; and Andy Patrizo, "The AI revolution has spawned a new chips arms race," *Ars Technica*, July 9, 2018. https://arstechnica.com/gadgets/2018/07/the-ai-revolution-has-spawned-a-new-chips-arms-race.

52. See, e.g., Ben Blan, "US-China trade war prompts rethink on supply chains," *Financial Times*, Sept. 3, 2018. https://www.ft.com/content/03e4f016-aa9a-11e8-94bd-cba20d67390c.

53. Hwang, pp. 29-32. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3147971.

54. Varian, p. 5. http://www.nber.org/chapters/c14017.pdf.

This becomes even more important as AI hardware becomes more specialized. Requiring a startup to buy different chips for the various life cycles of training and operating an ML algorithm would be a significant financial outlay and almost certainly hurt the ability of startups to compete. Fortunately, both AWS and Google Cloud have been competing with one another by adding specialized AI hardware as a part of their platform offerings.[55] This essentially allows startups to spread out the increased fixed costs of specialized hardware over a longer time horizon, which makes it more manageable.

In addition to the physical servers themselves, cloud computing companies are increasingly offering ML services like voice recognition, translation and image recognition to save startups the hassle of building their own software tools for each discrete task.[56] Again, it is difficult to understate how much easier this makes the process of launching a startup and it is a very positive development for the overall health of the AI ecosystem.[57]

This is closely related to the trend we have seen in the usage of distributed and open-software platforms like TensorFlow and GitHub, which provide ML platforms for startups to build, train and publish their software. While not hardware in the traditional sense, all of these can be thought of as a type of toolkit that exists around and supports the creation and development of AI. It is also noteworthy that many of these tools and platforms are effectively being developed and maintained for free by current incumbents.[58]

As this portion of the ecosystem largely seems to be developing in a healthy manner, the United States should be careful to avoid data-localization laws, excessive privacy laws, and other legislative efforts that might disrupt the careful balance.[59] On the whole, recommendations for this area should largely follow the Hippocratic Oath and "First, do no harm."

55. See, e.g., Cade Metz, "Google Makes Its Special A.I. Chips Available to Others," *The New York Times*, Feb. 12, 2018. https://www.nytimes.com/2018/02/12/technology/google-artificial-intelligence-chips.html.

56. Varian, p. 5. http://www.nber.org/chapters/c14017.pdf.

57. See, e.g., Kenji Kushida et al., "Diffusing the Cloud: Cloud Computing and Implications for Public Policy," *Journal of Industry, Competition and Trade* 11:3 (September 2011), pp 209-37. http://brie.berkeley.edu/publications/WP_197%20update%206.13.11.pdf.

58. While Google and Microsoft obviously benefit from the close developer relationships they maintain by offering TensorFlow and GitHub, it would be very difficult to argue the net effect of their existence is not pro-competitive.

59. See, e.g., Nigel Cory, "Cross-Border Data Flows: Where Are the Barriers, and What Do They Cost?", Information Technology and Innovation Foundation, May 1, 2017. https://itif.org/publications/2017/05/01/cross-border-data-flows-where-are-barriers-and-what-do-they-cost.

ANTITRUST CONSIDERATIONS

It is worth contrasting this general approach of reducing barriers to entry with another commonly cited remedy: stronger antitrust enforcement.[60] While concern over the level of domestic competition faced by GAFA is, of course, not unique to AI, it has certainly raised the stakes given how central the technology is to their current and future business models.

However, traditional antitrust measures may prove to be fairly difficult to implement and high risk for dealing with this perceived problem. After all, there are many plausible arguments supporting the current consolidated structure of the AI industry, particularly those that emphasize the importance of cross-cutting technical expertise, and the ability to leverage data and services from one business application to another.[61] While a full analysis of the antitrust implications of the AI industry is outside the scope of this paper, it is helpful to foreground the risks associated with such an approach.

If critics are right, breaking up or actively restricting the merger activities of large tech firms could lead to more innovation in the long run.[62] If these companies are indeed leveraging their significant market power to make it harder for startups to compete with them, breaking them up or constraining them could be a remedy.[63]

However, if critics are wrong about the optimal market structure of AI development and strong antitrust action is pursued, the consequences could be dire.[64] An increasing amount of evidence suggests that a small sliver of firms on the technological frontier have been responsible for the lion's share of productivity gains in the economy.[65] For this reason, breaking them up potentially risks killing the goose that lays the golden egg.[66]

By contrast, focusing on lower barriers to entry is a fairly low-risk strategy for injecting more competition into the AI landscape. If the United States can make it easier for startups to compete against large, established incumbents, it increases the likelihood of achieving the boosts to dynamism and innovation that antitrust advocates champion. Further, it would do so without risking the destruction of the current market equilibrium that is producing significant gains for consumers and for the broader economy. If GAFA can withstand the Schumpeterian winds67 of increased competition from startups, it is all the better for them.

However, as this paper has documented, there are significant barriers to entry in AI development that have boosted the market power of incumbent firms. If, in the absence of these barriers, new startups can successfully compete, it will be a win for innovation, consumers and for the dynamism of the economy as a whole.

One reason this strategy is lower risk than traditional antitrust remedies is because it does not impose a specific vision of market efficiency from the top down. Rather, it increases the level of competition from the bottom up in the hopes of displacing incumbent firms, if—and only if—the new firms are indeed more productive.

Furthermore, even if the current market structure is the most efficient, reducing the identified barriers to entry will increase the overall rate of innovation in the market by allowing AI to be developed more quickly. This will also aid in the diffusion of AI application across the rest of the economy, spreading the significant productivity gains that can result. Finally, it will make the United States more competitive on the international stage, as we compete with other nations to establish ourselves as the best place to develop and deploy AI systems.[68]

Considering the stakes involved and the relatively low risk associated with reducing barriers to entry, policymakers would be wise to focus on this agenda before moving on to more heavy-handed and higher-risk alternatives. Even in the

60. See, e.g., Robert Wright, "Google Must Be Stopped Before It Becomes An AI Monopoly," *Wired*, Feb. 23, 2018. https://www.wired.com/story/google-artificial-intelligence-monopoly.

61. See, e.g., Will Rinehart, "Breaking Up Tech Companies Means Breaking Up Teams And The Underlying Technology," American Action Forum, July 23, 2018. https://www.americanactionforum.org/insight/breaking-up-tech-means-breaking-up-technology-and-teams.

62. Editorial Board, "Break Up Google," *The Boston Globe*, June 14, 2018. https://apps.bostonglobe.com/opinion/graphics/2018/06/break-google.

63. Lina Khan, "Amazon's Antitrust Paradox," *Yale Law Journal* 26:3 (2016), pp. 710-805. http://digitalcommons.law.yale.edu/ylj/vol126/iss3/3.

64. Will Rinehart, "Breaking Up Big Tech Is Hard to Do," *The Wall Street Journal*, July 22, 2018. https://www.wsj.com/articles/breaking-up-big-tech-is-hard-to-do-1532290123.

65. See, e.g., Dan Andrews et al., "Frontier Firms, Technology Diffusion and Public Policy: Micro Evidence From OECD Countries," Organization for Economic Co-operation and Development, 2015. https://www.oecd.org/eco/growth/Frontier-Firms-Technology-Diffusion-and-Public-Policy-Micro-Evidence-from-OECD-Countries.pdf. Also, note that these large tech companies are by far the largest spenders on research and development in the United States. See, e.g., Rani Molla, "Tech companies spend more on R&D than any other companies in the U.S.," *Recode*, Sept. 1, 2017. https://www.recode.net/2017/9/1/16236506/tech-amazon-apple-gdp-spending-productivity.

66. For more on the high-risk nature of traditional antitrust enforcement in this sector, see Geoffrey Manne and Joshua Wright, "Innovation and the Limits of Antitrust," *Journal of Competition Law and Economics* 6:1 (2010), pp. 153-202. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1578762.

67. The economist Joseph Schumpeter popularized the term "creative destruction" and describes the effect of competition as feeling like a "gale" that "incessantly revolutionizes the economic structure from within, incessantly destroying the old one, incessantly creating a new one." Joseph Schumpeter, *Capitalism, Socialism and Democracy* (Routledge, 1942), pp. 82–83.

68. For more on the importance of international competition in AI, see, e.g., Michael Horowitz et al., "Strategic Competition in an Era of Artificial Intelligence," Center for New American Security, July 25, 2018. https://www.cnas.org/publications/reports/strategic-competition-in-an-era-of-artificial-intelligence.

event that strong antitrust enforcement is eventually called upon, enabling a more competitive ecosystem beforehand could help reduce the scope of the problem.

CONCLUSION

Artificial intelligence holds tremendous opportunity for our economy and for consumer benefit. However, the current barriers to entry in acquiring skilled talent and high-quality datasets may be impacting the number of startups that are able to compete successfully. And while the market for AI hardware appears to be developing in a healthy manner so far, policymakers should be careful not to implement policies that could backfire. To ensure a competitive and innovative ecosystem going forward, then, policymakers should prioritize reducing the barriers to entry as our first line of defense.

ABOUT THE AUTHOR

Caleb Watney is a technology policy fellow at the R Street Institute, where he leads the organization's work on emerging technologies, including autonomous vehicles, artificial intelligence, drones and robotics. He received his master's degree in economics from George Mason University.