

Research



A Framework for Increasing Competition and Diffusion in Artificial Intelligence

CALEB WATNEY, R STREET INSTITUTE | JANUARY 14, 2019

Artificial Intelligence (A.I.) is developing rapidly, and countries from around the globe are [beginning to articulate](#) national strategies for handling the political ramifications. With A.I. powering innovations such as driverless cars, autonomous drones, full-sequence genetic analytics, and powerful voice-assistant technology, the future [certainly looks](#) full of potential. Unsettled questions, however, 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, and [concerns](#) about lagging U.S. productivity and the slow pace of A.I. diffusion on the other, there is an underexamined overlap that connects these questions to the same set of policies: high barriers to entry due to supply-side constraints.

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

SUPPLY OF SKILLED A.I. ANALYSTS

Perhaps the single biggest bottleneck in A.I. development and application today is the supply of skilled data scientists and machine-learning engineers. Typical A.I. 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. In addition to these ballooning salaries, industry experts such as [Hal Varian](#) have pointed to the scarcity of adequate A.I. talent as the largest factor behind slow application in the economy.

Reform Our Immigration System to Allow More High-Skill A.I. 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, reforming 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 percent) were international. This influx of talent represents a significant portion of the overall A.I. 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. Due to a limited number of visa slots, however, only a fraction of these students [are allowed](#) to work in the country long term.

The primary pathway for these highly skilled immigrants to stay in the country is through the H-1B visa

program. For the past 16 years, however, 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 available spots. This discrepancy 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, and the paperwork burdens appear to be increasing. Both the financial and bureaucratic costs are easier for established firms to bear, given their larger size and increased resources.

In turn, this cost 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 — which would mean that the entire visa application process would have to be approached anew.

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 A.I. In particular, this reform 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 A.I. Talent

In addition to reforming our immigration pathways for high-skilled A.I. talent, it would be wise for the United States to extend more effort toward building up domestic talent. One way to achieve this end would be to better align incentives for companies to develop A.I. talent internally.

As the number of newly minted machine learning (ML) Ph.D. students continues to dwindle, some companies are looking at training employees internally to essentially create new supply. Such training, however, requires significant investment on the company's part, both in time and resources, to train new A.I. specialists this way, and the gains from this training are mostly captured by the newly trained worker in the form of higher wages. 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. 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. 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 A.I. workforce that is needed to fuel our economy. Given the pre-existing level of interest by employers in this strategy, it seems likely this could become a fruitful part of our domestic A.I. pipeline, if given more support.

SUPPLY OF DATA

In many ways, the supply of high-quality machine-readable training data is the key enabler of ML. Without access to some underexplored dataset, a team of talented A.I. specialists can be left twiddling their thumbs. Consumer data in the United States is particularly valuable, but here again large incumbents have significant (though not unsurmountable) [data advantages](#) over startups.

But 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 A.I. talent, this reform 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.

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](#) 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.

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.

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 the Health Insurance Portability and Accountability Act (HIPAA) has long been considered a [barrier to the development](#) of data sharing between medical professionals and companies. 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.

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 A.I. to diffuse throughout the economy, we must make sure the underlying data is [accessible](#).

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. Such an endeavor, however, 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 uncertainty, it is more likely that such a startup would avoid this potential legal minefield and consider what other datasets might be available with less risk.

Such is the ambiguous state of copyright enforcement in ML today. And it may also have important and underexplored applications for the state of competition in A.I.

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 reform could represent, then, a significant lever to create new arbitrage opportunities for scrappy startups willing to find and leverage interesting datasets.

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 [A.I. space](#).

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 A.I. 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 A.I. development.

While a natural and necessary part of the A.I. development process, such a trend toward specialized hardware does increase the fixed costs required to be competitive. This cost 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 A.I. analysts or datasets, maintaining access to valuable A.I. hardware is a key policy consideration.

Avoid Political Instability in International Supply Chains

As A.I. 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 that are vital to the supply chain, much of the production for particular parts (like back-end semiconductor fabrication) have 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.

To ensure access in spite of political tensions, large companies such as Apple, Google, and Nvidia are [beginning to re-shore](#) production of especially valuable chips. Smaller competitors and startups, however, 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 instability will have on smaller firms should be recognized. Ultimately, new foundries and semiconductor manufacturing plants will shift wherever they are most profitable. Accordingly, in the event of a long-term trade war, production could eventually shift elsewhere. Trade tensions, however, will certainly shape

short- and medium-term access to specialized hardware.

Maintain a Healthy Ecosystem Around Distributed Platforms

The other significant trend in A.I. hardware utilization is the growth of cloud-computing platforms such as 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. 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.

This becomes even more important as A.I. 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 A.I. hardware as a part of their platform offerings. This offering 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 such as voice recognition, translation, and image recognition to save startups the hassle of [building](#) their own software tools for each discrete task. 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 A.I. ecosystem.

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. On the whole, recommendations for this area should largely follow the Hippocratic Oath and “First, do no harm.”

WHAT ABOUT ANTITRUST?

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

Traditional antitrust measures, however, may prove to be both 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 A.I. 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.

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. 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](#).

If critics are wrong about the optimal market structure of A.I. development and strong antitrust action is pursued, however, the [consequences](#) could be dire. 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. Breaking up these large tech firms potentially risks [killing the goose](#) that lays the golden egg.

By contrast, focusing on lower barriers to entry is a fairly low-risk strategy for injecting more competition into the A.I. 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 incumbents can withstand the Schumpeterian winds of increased competition from startups, it is all the better for them.

As this essay suggests, there are significant barriers to entry in A.I. 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. To ensure a competitive and innovative ecosystem going forward, policymakers should prioritize reducing the barriers to entry as our first line of defense.