Digital Work in the Planetary Market

Edited by: Mark Graham, Fabian Ferrari

Citation:

Digital Work in the Planetary Market
Edited by: Mark Graham, Fabian Ferrari
DOI: 10.7551/mitpress/13835.001.0001
ISBN (electronic): 9780262369824
Publisher: The MIT Press
Published: 2022

The MIT Press

14 Data, Compute, Labor

Nick Srnicek

Pivotal to the future of the political economy of artificial intelligence (AI)[1]—and indeed, the future of current technology giants—is the question of whether AI is a centralizing technology. When we turn to the existing research on AI's impact on the economy, however, nearly all the attention has been on what we might call the automation/productivity channel, with discussion centered around whether, when, and how the spread of machine learning will automate and/or augment existing jobs (Royal Society and British Academy 2018; Frank et al. 2019; Acemoglu and Restrepo 2020). In this reading, AI is simply another labor-saving/-augmenting technology in a long line of such technologies. Much less attention has been given to how the nature of AI today may facilitate the centralization and concentration of capital, but this neglect has important consequences. For instance, one of the common arguments made by the defenders of today's technology giants is that their monopoly power is more precarious than it appears because of the ever-present threat of a disruptive innovation (Evans 2017; Christensen 2016; Pleatsikas and Teece 2001): IBM's mainframe monopoly lost out to personal computers; Microsoft's personal computing powerhouse lost out to mobile; and today's monopolies will eventually see similar disruption. Yet if the nature of the next major technologies is, for example, capital-intensive, and they have high barriers to entry, then we have good reason to believe disruption is unlikely. An understanding of the political economy of AI is therefore essential for understanding the stability of the current balance of power—and also for determining how the tech giants are consolidating power and acting strategically today. This knowledge is also crucial to our understanding of how the centralization of capital will play out across the planetary economy as American and Chinese platforms expand across the planet. In response to this gap in the literature, this chapter will examine the question of AI centralization in light of three key inputs: data, compute, and labor. Each of these inputs offers important insights into the global political economy of AI and its future trajectories.

The chapter will first examine the industry structure of AI. Far too often, the focus lies on the firms that *use* AI as opposed to the firms that *provide* AI. The latter, I will argue, are more important to understanding the nature of AI's political economy. The second section will show that most research on AI monopolies has been on data as an input into the production process, but in the third section, I will set out a schematic model of the AI production process that shows data is only one small part of a larger set of inputs and tasks.[2] The remainder of the chapter will then look at three key inputs—data, compute, and labor—discussing in turn why data is becoming less competitively important, and why compute and labor are becoming more significant. In the conclusion, I will attempt to draw out some initial geo-economic consequences of this new perspective on AI and monopolies.

Industry Structure

Within the last few years, each of the world's top technology companies (all based out of the US and China) has begun focusing on AI. Google, for instance, now declares itself to be an "AI-first company" and in May 2018 renamed Google Research as Google AI (Howard 2018). In March 2018, Microsoft reorganized its entire business and placed AI and the cloud together in their own unit (Nadella 2018). Baidu believes that its "strategic future relies on AI" (Clark 2017). And even more traditional companies like Apple and IBM are rapidly shifting to try to become AI-first companies. In contrast to the era of mutually exclusive fiefdoms, with platform giants dominating over their own particular areas, this new era presents an increasing convergence of the major platforms. The result is that companies—particularly Alibaba, Amazon, Baidu, Facebook, Google, Microsoft, and Tencent—are more frequently bumping up against each other. They are no longer just monopolists over particular industries and services, but active oligopolists fighting over an emerging AI and cloud computing sector.

Why are these companies moving into this field? Simply put, the belief (importantly, it remains a belief) is that AI provision will be an immensely lucrative and profitable field that will underpin the future of the global economy. In Marxist terms, AI is on its way to becoming a "general condition of production" (Dyer-Witheford, Kjosen, and Steinhoff 2019, 46–49)—or what mainstream economists call a general-purpose technology (Bresnahan and Trajtenberg 1995; Jovanovic and Rousseau 2005; Trajtenberg 2018). These are technologies that are not limited to particular sectors of the economy but instead have major impacts on production processes across the economy. Already we see signs of this, with a number of tasks increasingly being taken up by machine learning—from medical diagnoses to fraud detection, recommendation systems to translation services, demand forecasting to logistics optimization, and many more. Many of the tasks achievable by contemporary AI are common across businesses, and therefore the providers of AI services have large potential markets to tap into. Control over AI provision therefore means control over a new, and possibly immense, global economy–spanning technology worth perhaps trillions of dollars (Rao and Verweij 2017; Bughin et al. 2018).

It is not just the largest tech platforms that are getting involved in the AI industry, though. Billions of dollars are already being invested in AI research and deployment by venture capital, private equity, tech companies, and nontech companies (Bughin and Hazan 2017). Numerous countries have launched AI strategy documents with the aim of supporting and investing in nascent AI firms (Niklas and Dencik 2020), and venture capital investment and start-ups have been growing rapidly over the past decade (Furman and Seamans 2018, 5–6). Crunchbase, the database of start-ups, lists over 20,000 companies involved in AI as of August 2021.[3] We can bring some clarity to this rapid proliferation of activity by modeling the AI industry as comprising three distinct actors: AI providers, AI start-ups, and AI consumers.

AI providers involve the ownership and provision of the means of production for AI. This is currently the preserve and focus of the major platforms, with many of them adopting a cloud computing approach, or what is sometimes called "AI as a service." These companies (Alibaba, Amazon, Google, Microsoft, and Tencent, most notably) supply hardware, software, and even data for other companies to use. Rather than building their own internal AI models, the vast majority of companies are instead relying on this handful of AI providers to create, host, and maintain their AI models for them. The AI providers now offer increasingly industrialized versions of AI: standardized, off-the-shelf, and widely applicable models for tasks like image recognition, voice recognition, and natural language processing (Varian 2018, 7; Evans 2019; Clark 2020). In cases where AI models need more customization—for example, idiosyncratic data that the system needs to be trained on—these AI providers again offer the tools and computing power other companies need to perform these tasks. The result is that AI is effectively becoming like a utility, whether in the form of preexisting systems and services or in the form of the tools necessary to build their own, that firms can pay a fee (per inference or even per second) to access.

On a second level, there are the AI start-ups: companies that are creating new, artisanal AI services—often pitched at more niche markets than those the generic AI providers pitch to and often reliant on proprietary data. These companies are typically focused on specific problems that the big platforms are not in a position to recognize (e.g., automating the intricacies of a particular production process), and they tend to create specialized machine learning–based applications to provide solutions. These start-ups disprove the idea that the major platforms have "all the data," as they are often built on collecting data that others have ignored (Evans 2018). Google, for instance, may have a lot of search data, but it does not (yet) have telemetry data from wind turbines. But the AI start-ups that might collect such data still usually remain dependent on the AI providers: the storage capacity, the hardware, the database back ends, and the pretrained AI models that can then be customized are all rented from the major platforms. Yet these services are often quite expensive, with model training costing potentially US$100,000+ and with retraining often being necessary because of "data drift" (Casado and Bornstein 2020). The result is that the AI providers are routinely skimming around a quarter of the revenues of the AI start-ups, leaving the latter with small margins (Casado and Bornstein 2020).

Lastly, there is that group of firms that could be considered AI consumers: those who purchase and use AI services from others. This group includes, at its upper limits, the rest of the economy. As competitive pressures drive the adoption of machine learning across the economy, ever more firms will become AI consumers dependent upon a handful of AI providers (with some mediation by AI start-ups). Research on this group has been the primary focus of management literature on AI, as businesses seek to capture more value through the use of the technology (Agrawal, Gans, and Goldfarb 2018b; Davenport 2019).

Data and Monopoly

As I noted earlier, few scholars have to date examined the monopolization aspects of AI. Yet this monopolization tendency is particularly important insofar as all of the top AI companies are based out of the US or China. Even Europe remains far behind in terms of amounts invested, data collected, start-up successes, intellectual properties patented, and labor available. With this vast global disparity between the two leading countries and everyone else, the value created economy-wide by AI is more likely than not to be captured predominantly by the monopolists.

Among those who have looked at the potential for AI to create monopolization tendencies, two core arguments have emerged. First, there are those who have focused on the ways in which the *use* of AI will lead a handful of AI consumers to capture larger and larger market shares. The McKinsey Global Institute, for instance, argues that those who more rapidly adopt AI will gain significantly in the coming decades, while those who do not will fall further and further behind (Bughin et al. 2018, 39–41). These early adopters tend to already be highly digitized and large-scale, and therefore primed to take advantage of the benefits that AI might offer (Bughin and Manyika

2018). Research in management studies comes to similar conclusions: late adopters of AI will struggle because it takes a lot of time and effort to collect data and adapt AI systems to local conditions and requirements (Mahidhar and Davenport 2018). Those who get a head start on this process are deemed likely to pull away from their competitors (Mahidhar and Davenport 2018). More elaborate accounts have focused on how algorithms across a number of firms may interact in such a way as to unintentionally create collusion and an oligopolistic market structure (Ezrachi and Stucke 2016, 2015).

If this first series of arguments has applied to the firms that use AI (consumers), a second set of arguments applies to both the consumers and providers. Here, attention has centered on the role of data in the process of training AI models. This is particularly the case for deep learning, which relies upon massive amounts of data, and for which research has repeatedly found that more data makes for better AI (Sun et al. 2017). Given that the extraction of data is already concentrated in the hands of a few major platforms, it is hardly a leap to expect that a technology that relies on that data will be similarly concentrated. We might therefore expect the technological requirements of contemporary AI to induce a strong tendency toward centralization and concentration of capital. Yet the debate so far has been divided.

On the one hand, there are those who are largely skeptical about the possibility of machine learning increasing monopolization of providers. Hal Varian, a well-paid employee of Google, is the clearest articulator of this position. In his account, AI is unlikely to involve increasing returns to scale on the supply side, as the cost of supplying AI software does not significantly decrease after the creation of the software. Whereas traditional software could replicate and sell innumerable copies at decreasing marginal cost, today's AI systems instead require updates and other continual improvements that mean costs continue (Varian 2018, 16). Neither is AI likely to have demand-side increasing returns to scale, as Varian believes network effects will not operate here—and in the cases where they might (e.g., firms choosing an AI provider because it is well-known), these are no different than what happens in any other industry. Lock-in, or path dependency, is deemed not to be a major issue either, as developments like containerization[4] enable firms to shift from one AI provider to another (Varian 2018, 19). And lastly, data is argued to have decreasing returns to scale, in that while more data may mean more accurate predictions from models, it increasingly takes more and more data to eke out smaller and smaller predictive gains (Agrawal, Gans, and Goldfarb 2018a, 20–21).

Similar arguments against data's role in facilitating monopolization come from the venture capital firm Andreessen Horowitz. As they note, after a certain point, getting the data needed to improve an AI system can become systematically more difficult even as the benefit to the predictive accuracy becomes increasingly marginal (Casado and

Lauten 2019). New data often overlaps with existing data, the relevant data needed can be difficult to find and can involve edge cases, and since it involves edge cases, the number of times the data will be relevant can be minimal. The former chair of Barack Obama's Council of Economic Advisors likewise finds in a review of the literature that there is "limited evidence of increasing returns to scale for data" (Furman and Seamans 2018, 19).

On the other hand, there are those who argue that, at least under certain conditions, data does generate a virtuous cycle that can lead to a handful of firms dominating. Agrawal, Gans, and Goldfarb (2018a, 21), for instance, agree with Varian that more data may not have increasing returns for technical value but counter that it often does have increasing returns for economic value. For example, in many image recognition applications, marginally better accuracy may not matter. But in some circumstances (e.g., medical diagnoses), an algorithm that is accurate 99 percent of the time will be vastly more useful than one that is accurate only 95 percent of the time. In such cases, the decreasing returns of data still manifest significant economic advantages. Monopolization can emerge here, as more data will lead to better-quality services, which leads to more customers, which leads to more data—and so on (Casado and Lauten 2019). And perhaps the boldest version of the argument comes from another venture capitalist, Kai-Fu Lee, who is one of the few to also draw out the geopolitical implications of this tendency (Lee 2018). As he writes,

> First, most of the money being made from artificial intelligence will go to the United States and China. A.I. is an industry in which strength begets strength: The more data you have, the better your product; the better your product, the more data you can collect; the more data you can collect, the more talent you can attract; the more talent you can attract, the better your product. It's a virtuous circle, and the United States and China have already amassed the talent, market share and data to set it in motion. (Lee 2017)

Here we see that not only is data a facilitator of monopolization, but it will also facilitate a planetary concentration of AI's value in the hands of two beneficiary countries. All others, in Lee's account, will become dependent on the US and China for AI services.

These latter arguments about the significance of data and its potential to create virtuous cycles appear far more plausible than the denial of this trajectory. Regulators appear to have agreed, with much of the discussion around how to regulate AI companies (with respect to their political economy rather than their ethical applications) hinging upon proposals for data sharing. A recent European Commission report, for instance, suggests that "where specific circumstances so dictate, access to data should be made compulsory, where appropriate under fair, transparent, reasonable, proportionate and/or nondiscriminatory conditions" (European Commission 2020, 13). Similar policy proposals have been put forward in Germany as well (Nahles 2018).

AI Production Process

While the focus of existing research on monopolization and AI has been on who dominates in data collection, the AI production process involves several other stages. Broadly speaking, we can distinguish between four different stages (Dong 2017): data collection, data processing, model production, and model deployment/monitoring/retraining. The first of these, data collection, is the finding of data to feed into and train machine learning models.

The second stage is data processing, involving the cleaning up and (often) labeling of data. In terms of concrete labor time, this is the most involved part of the entire process; it typically requires scores of workers to accomplish. One study, for instance, found that this work took around 80 percent of the labor time needed to build an AI system (Cognilytica 2019). A raft of new companies has emerged to offer data labeling services, often relying on marginalized populations within countries (e.g., prison labor or refugees; see also chapter 9) or marginalized populations globally (e.g., low-wage workers in Africa, Asia, and Latin America) (Batha 2018; Cadell 2019; Chen 2019; Gray and Suri 2019; Metz 2019; Murgia 2019b; Anwar and Graham 2020; see also chapter 6). There is an emerging global inequality here, with some businesses in the Global North offering more specialized data processing services—and charging higher fees as a result. The production of properly labeled data for sensitive issues (e.g., legal classification, driverless car data, or medical imagery) requires people familiar with the subject area (Peng 2019). And even apparently low-level data labeling—such as facial key-point labeling—increasingly demands extraordinary skills: "The task was much simpler a few years ago, when labellers only had to put several dots on a human face. Now, facial key-point labelling can involve up to 206 dots—8+ on each eyebrow, 20+ on the lips, 17+ along the jawline, and so on" (Peng 2019). The need for high-quality labeling is leading to the emergence of an industry to provide it, while lower-quality labeling is outsourced to the margins of the planetary economy. The growth of the AI economy is following well-worn paths already set out by existing global hierarchies.

Once the data has been collected and labeled, the third stage of the AI production process is model production: the process of selecting an AI model and feeding the data into it in order to train it. This stage often requires highly skilled labor—to tune algorithms and other tasks—though it is much less labor-intensive than earlier stages. It is, however, computationally intensive—a point to which I will return later. Lastly, there is the deployment and monitoring stage. This involves seeing whether the model works in practice (e.g., does it produce significant bias problems?), as well as updating and retraining old models. The latter is particularly important, as models typically degrade over time due to the fundamental limitation that they are trained on a finite dataset (Abrahamson 2019). As the world changes, so too do the patterns that the model is aiming to mirror, and therefore this final stage of the AI process is often a matter of continually retraining models to reflect new data.

With this expanded image of the AI production process in hand, the argument I want to make in the remainder of this chapter is that the first two stages are decreasing sources of competitive advantage (and decreasing monopolistic footholds), while the latter two are becoming the real battleground for who will control AI.[5] Yet as we saw in the review of the existing literature, the latter two stages are largely ignored in favor of a focus on data alone.

Data

Let us deal with a first question that might come to mind. Despite all the attention paid to the importance of data and its being lauded as the new oil, why is data losing significance as a competitive advantage? The first reason is the general spread of the platform business model and its unique capacities to collect data (Srnicek 2016). Whereas a decade ago relatively few companies had platforms and systems in place to monitor and record data, today the data-centric business model has become increasingly widespread. Whether legacy companies or start-ups, everyone is trying to collect their own proprietary data. There are, of course, still significant disparities in the amount of data being collected by various firms. But the gap in the amount of data being collected between the companies collecting vast amounts and the companies collecting none is arguably decreasing.

A second reason why data is losing some of its competitive advantage is the explosive growth of open datasets. For instance, Google subsidiary Waymo's Open Dataset for training driverless vehicles contains nearly 17 hours of video, with labeling for 22 million 2D objects and 25 million 3D objects (Peng 2019). Google's Open Images Dataset contains 9.2 million photos, with over 30 million labels for almost 20,000 concepts (Kuznetsova et al. 2020). And data is available for even the most niche of interests. Clemson University and the University of Essex, for example, released a database of 4.5 million transcriptions of speeches given in the Irish parliament between 1919 and 2013 (Herzog and Mikhaylov 2017). As ever more vast and labeled datasets are being made available, new companies will have less need to go out and find their own data. The problem of how to bootstrap from no data is, simply put, increasingly less of a problem. To be sure, none of this is to say that the open datasets are equivalent in value to proprietary ones—but they do go some way toward lessening the competitive advantages of the latter and reducing the problems of bootstrapping training from nothing.

The final, and potentially most significant, reason for data's relative decline is the emergence of synthetic data. Rather than relying on getting data from the real world, a growing number of companies is synthetically creating new data and synthetically augmenting low-quality and sparse data. The perhaps best-known examples of this involve generative adversarial networks (GANs) creating realistic-looking photos of individuals. But others have used GANs to generate videos that algorithms can then be trained on. And other researchers have taken small datasets (of plants in this case) and used GANs to create more images. After comparing systems trained on the two datasets, the researchers found that the larger (part synthetic) dataset ended up having less overfitting and slightly more accuracy (Giuffrida, Scharr, and Tsaftaris 2017). Other approaches to synthetic data involve artificial environments—often artificial worlds created with videogame engines—that AI agents interact with and learn from. The explosion of videogame-playing AI, often based on reinforcement learning, is one expression of this. Artificial environments offer several advantages—if real-world experiments with AI-driven robots take significant time, running the same experiments in simulations can be vastly quicker (and safer). In this vein, even well-known videogames like the Grand Theft Auto series are being deployed as environments to train driverless cars (Li et al. 2017). These environments are also becoming industrialized and open-source, with OpenAI Universe including over 1,000 environments and Facebook releasing the TorchCraft environment for researchers to use.

These processes of creating data from nothing are allowing a number of start-ups lacking their own data to get past the initial hurdles (Simonite 2018) and have led some to argue that synthetic data will lead to the end of any competitive advantage in the area of data for the tech giants (Nisselson 2018). But incumbent firms are also using these methods, with, for example, Amazon using GANs to create e-commerce data, Facebook generating synthetic users to test out its platform, and Google creating synthetic skin lesion images to train healthcare AI (Kumar, Biswas, and Sanyal 2018; Ahlgren et al. 2020; Kohlberger and Liu 2020). The result of all this may be AI companies moving from "data competition" to "environment competition" (Clark 2018a). As these new approaches to AI continue to gain traction, it is more likely that data will become less of a competitive advantage.

Compute

If data collection and labeling are becoming less significant, the other stages of the production process are increasingly where AI monopolies and moats are being built. This is for two reasons in particular: the concentrated ownership of immense computing resources (compute) and the systems and lures built for attracting the small supply of high-skill workers.[6]

With respect to the former, increases in computing power have been driving advances in AI—not only during its most recent deep learning incarnation but for many decades now (Sutton 2019). In the past decade, we have seen AI models getting larger as well, with more parameters than ever before. In 2019, for example, NVIDIA released a model with 8.3 billion parameters, while Google released a model with over 50 billion parameters (Bapna and Firat 2019; Toole 2019). To meet the immense challenges of training these models, AI hardware has been scaling up to data centers and supercomputers. More attention is also being given to how to network hundreds and even thousands of graphics processing units (GPUs) together in order to train these massive models (Hazelwood et al. 2018; Laanait et al. 2019). The result of all this has been a significant leap forward in the computing power being used to train the largest AI systems. Between 1959 and 2012, the use of compute for training AI systems increased at broadly the same rate as Moore's Law—doubling every 24 months (Amodei and Hernandez 2018). Yet between 2012 and 2018, there was a 300,000× increase in the amount of compute used to train the largest models—a doubling every 3.4 months. This rapid increase has been fostered by better chips and, more significantly, by an increase in the ability to use parallel processing (Amodei and Hernandez 2018). Beyond the production stage of models, the deployment stage is also ramping up in terms of compute requirements, making some of the largest models increasingly unwieldy (Kaiser 2020).

It is, unsurprisingly, the AI providers who are positioned to be able to use and deploy the sorts of compute needed for cutting-edge AI. The shift to cloud computing is an expression of this, as it is data center–scale computing that is required—and providers can use GPUs at one point to further their own research and at another point as a rental for a small AI start-up. Computing resources are, in turn, an expression of financial resources. For example, it cost an estimated $35 million to train AlphaGo Zero, the groundbreaking self-taught AI program for playing the game of Go (note that this does not include the cost of researchers or anything that went into initially building the project) (Huang 2018). Data centers are immensely expensive propositions for any company. And while detailed figures on the amounts being spent on this infrastructure are not available publicly, the financial statements of the big cloud companies all reveal tens of billions of dollars being poured annually into fixed capital. Amazon, Microsoft, and Google, for example, collectively had $73.5 billion in capital expenditures in 2019 (Fitzgerald 2020). Far from being immaterial companies, these are significantly embodied companies. And as these companies turn to designing their own specialized computer chips to gain more speed and power, the entry fees to compete with them are growing. Nearly all the major tech companies are investing in their own designs (e.g., Google's Tensor Processing Units) or buying up smaller chip start-ups (e.g., Amazon's purchase of Annapurna Labs). The economics of all this again favor the largest tech companies with the capital to invest in new chip design, buy the chips in data center quantities, and deploy them for their own benefit. The capital expenditure required for this scale of computing creates extremely high barriers to entry, the result being that AI provision is ultimately a market in which only a handful of companies globally stand a chance.

This scale of compute lends itself to further benefits for these companies. First, AI systems tend to perform better when they have more compute. As one review of the impact of compute notes, "There is a close tie from compute operation rate (e.g., floating point operations, or 'FLOPs') to model accuracy improvements" (Hestness et al. 2017, 13). And while there are diminishing returns to the value of increasing available compute, as we saw earlier, significant economic value can still be extracted from even marginal increases in accuracy.

More compute also enables companies to train and retrain models much more quickly than their competitors. As AI remains an empirical science, it involves running a number of experiments to see what works best—tuning hyperparameters, testing on data from outside the training set, debugging any problems, and so on. The more rapidly a firm can do this, the more rapidly it can deploy models to users. Moreover, as the world changes, models degrade and need updating. Again, the more rapidly one can retrain models, the better they will perform and the more users they are likely to attract. The differentials in speed between an average firm and a major platform can be immense. For example, in 2012, to train a model on the ImageNet dataset to a 75 percent degree of accuracy took 7–14 days with a single GPU. By 2019, one Chinese company had managed to train a model on the same dataset, to the same degree of accuracy, in 75 seconds with 2048 GPUs—a time reduction of 99.9 percent (Clark 2019). These sorts of advantages can be massive in a rapidly changing world.

Lastly, more computing power enables better research. It allows researchers to try ideas at scales that are unavailable to smaller firms. Innovations, as a result, are more likely to come from the larger AI providers. These innovations may eventually filter down to smaller firms as improvements and efficiencies make them more readily available. But as one investor puts it, "Having a really, really big computer is kind of like a time warp, in that you can do things that aren't economical now but will be economically [feasible] maybe a decade from now" (Levy 2017). More computing power also lets researchers explore the boundaries and limits of different approaches in much more thorough ways—determining, in one example, whether reinforcement learning or evolutionary learning is better for a particular problem set (Clark 2018b). In another instance, Facebook researchers used large-scale training to find faster ways to do machine translation (Edunov, Auli, and Ott 2018). Additionally, the rapidity of training lets these researchers determine what are unfruitful ventures and what might be productive avenues far more quickly than those who must limit themselves to testing on smaller systems that may not scale up (Clark 2018c). More compute, in the end, allows the big AI providers to more rapidly and effectively research AI possibilities and gain even more ground on their competitors.

Labor

To make effective use of compute resources, though, requires workers with the skills to put these systems together in the first place. The task of scaling from one GPU to 1,000 is highly technical and challenging, and workers capable of doing it remain in short supply. This leads us to the third key element of monopolization in the AI world: labor.

On the first level, the short supply of these workers means they can often command very large salaries. Average salaries for data scientists now range into six figures in America. DeepMind, arguably the world's leading AI center, spent nearly £400 million on "staff and related costs" in 2018.[7] This means that companies with the cash to spend on the best talent are the ones who tend to be winning the race for AI talent. Academia has been one of the big losers, with a major brain drain of AI researchers from universities to companies. Large salaries and access to unprecedented amounts of compute are drawing these researchers into the arms of the top AI companies (Murgia 2019a).

Yet beneath the flashy salaries lies a series of much subtler ways in which the AI providers are channeling talent their way. In particular, the seemingly noncapitalist practice of releasing their AI software for free in fact obscures a significant capitalist battle between the major companies. These open-source frameworks offer a range of premade tools, libraries, and interfaces for others to build their AI models with—often based on the same tools that companies are using internally. At present, Google's TensorFlow is the most popular in the industry, though others have more niche markets, and Facebook's PyTorch is coming to dominate the research sector (He 2019; Kaggle 2019, 19). Why would these companies give away such potentially valuable software, though? First, these projects invoke what Paolo Virno once called the "communism of capital" (Virno 2004, 110).[8] Such open-source projects foster and support communities of labor that provide inputs back into the software—all for free. The more widely a framework is used, the more likely it is that the community will find bugs, add features, and generally innovate and develop the software in useful ways. And the easiest way to ensure widespread usage is through lowering the cost barrier by open-sourcing these projects.

In addition, a successful framework builds up a community of developers who know how to work within a particular company. Graduate students, for instance, who have trained with these tools are primed to slide into a corporate position when they leave university looking for a career. Frameworks become feeder networks for the emerging generations of talent. Such is the importance of these frameworks that other major companies have begun working together in an attempt to compete against Google's TensorFlow. Apple and Amazon, for instance, have teamed up to enable applications written in Amazon's MXNet framework to be easily translated into Apple's Core ML framework (Menant and Gupta 2017). Microsoft and Facebook, meanwhile, created the Open Neural Network Exchange (ONNX) format, which makes their own frameworks (CNTK, PyTorch, and Caffe2) interoperable (Boyd 2017). Each of these strategic capitalist alliances is an effort to overthrow the current dominance of TensorFlow and the advantages that it lends Google. In any case, though, with the channeling of labor away from lower-tier companies that the communism of capital entails, all the frameworks are increasingly controlled by the top-tier AI companies. The powerful grow stronger.

Conclusion

The argument I have tried to set out here is that contemporary AI is a monopolizing technology but that the often-assumed driver of this tendency—data—is less significant than believed. Instead, contemporary AI is increasingly driven by the inputs of compute and labor, and these are forming the real competitive advantages for the largest AI providers as they continue to pull away from any possible challengers. Far from being a disruptive threat to existing technology giants, AI appears set to further consolidate their power. In this conclusion, I will briefly examine some of the important implications that this analysis might have for uneven planetary economic development.

If data, for instance, were the only key input to the AI production process, one could imagine something like a national data commons being sufficient to chart a path for digital sovereignty. Take control over data, and you would have taken control of the key resource that gives the AI providers their power. This belief, as we saw earlier, seems to motivate a number of the policy proposals currently being put forth across Europe. However, if compute and labor are also key aspects, it is hard to avoid the conclusion that the biggest cloud AI companies, centered solely in the US and China, will continue to pull away from the rest regardless of data policies. Other countries and other companies have not shown an ability to invest the same amount into fixed capital as the top American AI firms, nor do they have much capacity to retain talent when a company like Google is willing to pay enormous salaries and give luxurious amounts of research freedom to data scientists and other skilled workers.

These companies, moreover, are rapidly stretching their tentacles across the remainder of the world. Kai-Fu Lee (2017), for example, paints a plausible picture of US companies carving up the developed world while Chinese platforms expand across the developing world. In this possible future, it seems likely that much of the world, developed or otherwise, will remain relatively low in the AI value chain (Weber 2017). Low-waged data labeling is already spread across the peripheries, reliant on hyperexploited and marginalized workers. There is an unevenness here as well, as we saw earlier, with high-skill labeling being brought into the metropole of the AI world. In the world of AI consumers, other countries and the US will continue to be able to grab a part of the AI value chain. Start-ups can find novel uses of machine learning and apply them to new products. Yet they will remain tenants on the clouds provided by the biggest AI companies, dutifully paying their rents to these American and Chinese companies. Meanwhile, start-ups that appear promising are all too likely to be swallowed up by the tech giants. Activity in mergers and acquisitions related to AI, for instance, increased by 500 percent between 2013 and 2017. Between 2010 and 2019, Apple made over 20 AI acquisitions, Google made 14, and Microsoft made 10.[9] By comparison, the vast majority of the companies that purchased an AI company in that decade only bought a single company.[10] The largest companies continue to pull away, and the market for AI provision continues to consolidate.

This means that in thinking about digital development in the Global South, a focus on start-ups is insufficient to overcome existing imbalances. Moreover, the impacts on broader ideas of economic development are likely to be significant. Not only is the profit of the emerging global value chains for AI being captured by a handful of companies, but the secondary effects of that value capture—the conglomeration effects and other spillovers from AI growth—are also likely to be concentrated in a handful of countries (Weber 2017, 412). To put it starkly, it may turn out that while workers in Kampala are spending their poorly paid time labeling images of faces, wealth and talent are creating virtuous cycles of local growth in Silicon Valley and Shenzhen. Developing countries—and many developed countries—look likely to remain trapped in positions of relative digital underdevelopment. The emerging planetary value chain of AI is a profoundly unequal one.

Let me conclude with three points. First, as I have argued here, the monopolization tendency is not just—or even primarily—a data issue. Monopolization is driven more by the barriers to entry posed by fixed capital and the virtuous cycles that compute and labor are generating for the AI providers. The academic literature has, to date, largely neglected to examine these elements. Second, a consequence of the preceding argument is that open-source software is not an alternative so much as a strategic tool for these AI platforms. Existing arguments about how large tech companies freely build their proprietary empires on top of open-source software must be supplemented with attention to the ways in which free—and waged—labor is brought into the companies' ambit via things like open-source frameworks. Lastly, another notable consequence is that policy in response to AI development must go beyond the fascination with data. If, as I argue, hardware and labor are important inputs too, then opening up data is an ineffective idea at best and a counterproductive one at worst. It could simply mean that the tech giants get access to even more free data—while everyone else trains their open data on Amazon's servers. If we want to take back control over big tech, we need to pay attention to more than just data.

Notes

1. By "artificial intelligence," I specifically mean the constellation of machine learning models and techniques that have emerged in the wake of the 2012 deep learning revival inspired by the ImageNet success of Krizhevsky, Sutskever, and Hinton (2012).

2. A recent paper by Mucha and Seppälä (2020) has made clear just how narrow most research on the economics of AI currently is.

3. See https://www.crunchbase.com/search/organization.companies/6fc9f338b99a553e26331718a9377efc.

4. Containerization is a recent development that enables applications to run more easily in any cloud environment rather than being built for and able to run in only specific ones.

5. Thanks to Jack Clark's *Import AI* newsletter for initially bringing my attention to the significance of other aspects in contemporary AI development and competition. This chapter attempts to build on and systematize some of the arguments he has made in his newsletter.

6. *Compute* here is a term commonly used in the cloud computing and AI industries to refer to computing resources (as opposed to, say, network resources or memory resources).

7. See https://beta.companieshouse.gov.uk/company/07386350/filing-history.

8. Thanks to Nick Dyer-Witheford, Atle Mikkola Kjolsen, and James Steinhoff for reminding me of this reference.

9. We have not focused on Apple in this piece because their AI strategy is focused more on devices than on cloud platforms. While the focus on devices remains lucrative for Apple at the moment, AI companies based on cloud platforms appear to be far more significant in their implications.

10. See https://interactives.cbinsights.com/artificial-intelligence-acquisitions-by-famga/.

References

Abrahamson, Joseph. 2019. "Model Performance Often Degrades over Time." Simplicial, April 22.

Acemoglu, Daron, and Pascual Restrepo. 2020. "The Wrong Kind of AI? Artificial Intelligence and the Future of Labour Demand." *Cambridge Journal of Regions, Economy and Society* 13 (1): 25–35.

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. 2018a. "Economic Policy for Artificial Intelligence." Working Paper 24690. Cambridge, MA: National Bureau of Economic Research.

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. 2018b. *Prediction Machines: The Simple Economics of Artificial Intelligence*. Boston: Harvard Business Review Press.

Ahlgren, John, Maria Eugenia Berezin, Kinga Bojarczuk, Elena Dulskyte, Inna Dvortsova, Johann George, Natalija Gucevska, Mark Harman, Ralf Lämmel, Erik Meijer, Silvia Sapora, and Justin Spahr-Summers. 2020. "WES: Agent-Based User Interaction Simulation on Real Infrastructure." Facebook Research. https://research.fb.com/publications/wes-agent-based-user-interaction-simulation-on-real-infrastructure/.

Amodei, Dario, and Danny Hernandez. 2018. "AI and Compute." *OpenAI*, May 16. https://openai.com/blog/ai-and-compute/.

Anwar, Mohammad Amir, and Mark Graham. 2020. "Digital Labour at Economic Margins: African Workers and the Global Information Economy." *Review of African Political Economy* 47 (163): 95–105.

Bapna, Ankur, and Orhan Firat. 2019. "Exploring Massively Multilingual, Massive Neural Machine Translation." Google AI Blog, October 11.

Batha, Emma. 2018. "Mobile App Pays Refugees to Boost Artificial Intelligence." Reuters, November 15. https://www.reuters.com/article/us-refugees-conference-tech-jobs-idUSKCN1NK2SR.

Boyd, Eric. 2017. "Microsoft and Facebook Create Open Ecosystem for AI Model Interoperability." *Microsoft Azure* (blog), September 7. https://azure.microsoft.com/en-us/blog/microsoft-and-facebook-create-open-ecosystem-for-ai-model-interoperability/.

Bresnahan, Timothy F., and M. Trajtenberg. 1995. "General Purpose Technologies: 'Engines of Growth'?" *Journal of Econometrics* 65 (1): 83–108.

Bughin, Jacques, and Eric Hazan. 2017. "The New Spring of Artificial Intelligence: A Few Early Economies." *VoxEU*, August 21. https://voxeu.org/article/new-spring-artificial-intelligence-few-early-economics.

Bughin, Jacques, and James Manyika. 2018. "Technology Convergence and AI Divides: A Simulation Appraisal." *VoxEU*, September 7. https://voxeu.org/article/technology-convergence-and-ai-divides.

Bughin, Jacques, Jeongmin Seong, James Manyika, Michael Chui, and Raoul Joshi. 2018. "Notes from the AI Frontier: Modeling the Global Economic Impact of AI." Washington, DC: McKinsey Global Institute. https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy.

Cadell, Cate. 2019. “Faces for Cookware: Data Collection Industry Flourishes as China Pursues AI Ambitions.” Reuters, June 27. https://www.reuters.com/article/us-china-ai-data-insight-idUSKCN1TS3EA.

Casado, Martin, and Matt Bornstein. 2020. “The New Business of AI (and How It’s Different from Traditional Software).” Andreessen Horowitz, February 16. https://a16z.com/2020/02/16/the-new-business-of-ai-and-how-its-different-from-traditional-software/.

Casado, Martin, and Peter Lauten. 2019. “The Empty Promise of Data Moats.” Andreessen Horowitz, May 9. https://a16z.com/2019/05/09/data-network-effects-moats/.

Chen, Angela. 2019. “Inmates in Finland Are Training AI as Part of Prison Labor.” *The Verge*, March 28. https://www.theverge.com/2019/3/28/18285572/prison-labor-finland-artificial-intelligence-data-tagging-vainu.

Christensen, Clayton M. 2016. *The Innovator’s Dilemma: When New Technologies Cause Great Firms to Fail*. Boston: Harvard Business Review Press.

Clark, Jack. 2017. “The AI Data Grab, the Value of Simplicity, and a Technique for Automated Gardening.” *Import AI* no. 41, May 8. https://jack-clark.net/2017/05/08/.

Clark, Jack. 2018a. “Virtual Beijing with ParallelEye, NVIDIA Tweaks GPU Licensing, and Saving Money by Getting AI to Help Humans Label Data Generated by AI.” *Import AI* no. 75, January 1. https://jack-clark.net/2018/01/01/.

Clark, Jack. 2018b. “Trading Cryptocurrency with Deep Learning; Google Shows Why Evolutionary Methods Beat RL (for Now); and Using Iwatch Telemetry for AI Health Diagnosis.” *Import AI* no. 81, February 12. https://jack-clark.net/2018/02/12/.

Clark, Jack. 2018c. “Sony Researchers Make Ultra-Fast ImageNet Training Breakthrough; Berkeley Researchers Tackle StarCraft II with Modular RL System; and Germany Adds €3bn for AI Research.” *Import AI* no. 121, November 19. https://jack-clark.net/2018/11/19/.

Clark, Jack. 2019. “AIs Play Doom at Thousands of Frames per Second; NeurIPS Wants Reproducible Research; and Google Creates & Scraps AI Ethics Council.” *Import AI* no. 141, April 8. https://jack-clark.net/2019/04/08/.

Clark, Jack. 2020. “The Industrialization of AI, BERT Goes Dutch, plus, AI Metrics Consolidation.” *Import AI* no. 182, January 27. https://jack-clark.net/2020/01/27/.

Cognilytica. 2019. “Data Engineering, Preparation, and Labeling for AI 2019.” *Cognilytica Research*, January 31.

Davenport, Thomas. 2019. *The AI Advantage: How to Put the Artificial Intelligence Revolution to Work*. Cambridge, MA: MIT Press.

Dong, Catherine. 2017. “The Evolution of Machine Learning.” *TechCrunch*, August 8. http://social.techcrunch.com/2017/08/08/the-evolution-of-machine-learning/.

Dyer-Witheford, Nick, Atle Mikkola Kjosen, and James Steinhoff. 2019. *Inhuman Power: Artificial Intelligence and the Future of Capitalism*. London: Pluto Press.

Edunov, Sergey, Michael Auli, and Myle Ott. 2018. "Scaling Neural Machine Translation to Bigger Data Sets with Faster Training and Inference." Facebook Engineering, September 7. https://engineering.fb.com/ai-research/scaling-neural-machine-translation-to-bigger-data-sets-with-faster-training-and-inference/.

European Commission. 2020. "A European Strategy for Data." Brussels: European Commission. https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf.

Evans, Benedict. 2018. "Does AI Make Strong Tech Companies Stronger?" *Benedict Evans* (blog), December 20. https://www.ben-evans.com/benedictevans/2018/12/19/does-ai-make-strong-tech-companies-stronger.

Evans, Benedict. 2019. "The Deployment Phase of Machine Learning." *Benedict Evans* (blog), October 4. https://www.ben-evans.com/benedictevans/2019/10/4/machine-learning-deployment.

Evans, David S. 2017. "Why the Dynamics of Competition for Online Platforms Leads to Sleepless Nights but Not Sleepy Monopolies." *SSRN*, August 3. https://ssrn.com/abstract=3009438.

Ezrachi, Ariel, and Maurice E. Stucke. 2015. "Artificial Intelligence & Collusion: When Computers Inhibit Competition." *SSRN*, April 9. https://ssrn.com/abstract=2591874.

Ezrachi, Ariel, and Maurice E. Stucke. 2016. *Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy*. Cambridge, MA: Harvard University Press.

Fitzgerald, Charles. 2020. "Follow the CAPEX: Cloud Table Stakes 2019 Edition." *Platformonomics* (blog), February 11. https://www.platformonomics.com/2020/02/follow-the-capex-cloud-table-stakes-2019-edition/.

Frank, Morgan R., David Autor, James E. Bessen, Erik Brynjolfsson, Manuel Cebrian, David J. Deming, Maryann Feldman, Matthew Groh, José Lobo, Esteban Moro, Dashun Wang, Hyejin Youn, and Iyad Rahwan. 2019. "Toward Understanding the Impact of Artificial Intelligence on Labor." *Proceedings of the National Academy of Sciences* 116 (14): 6531–6539.

Furman, Jason, and Robert Seamans. 2018. "AI and the Economy." Working Paper 24689. Cambridge, MA: National Bureau of Economic Research.

Giuffrida, Mario Valerio, Hanno Scharr, and Sotirios A. Tsaftaris. 2017. "ARIGAN: Synthetic Arabidopsis Plants Using Generative Adversarial Network." Preprint, submitted September 4. http://arxiv.org/abs/1709.00938.

Gray, Mary L., and Siddharth Suri. 2019. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Boston: Houghton Mifflin Harcourt.

Hazelwood, Kim, Sarah Bird, David Brooks, Soumith Chintala, Utku Diril, Dmytro Dzhulgakov, Mohamed Fawzy, Bill Jia, Yangqing Jia, Aditya Kalro, James Law, Kevin Lee, Jason Lu, Pieter Noordhuis, Misha Smelyanskiy, Liang Xiong, and Xiaodong Wang. 2018. "Applied Machine Learning at Facebook: A Datacenter Infrastructure Perspective." In *2018 IEEE International Symposium on High*

Performance Computer Architecture (HPCA), 620–629. https://research.fb.com/wp-content/uploads/2017/12/hpca-2018-facebook.pdf.

He, Horace. 2019. "The State of Machine Learning Frameworks in 2019." *The Gradient*, October 10. https://thegradient.pub/state-of-ml-frameworks-2019-pytorch-dominates-research-tensorflow-dominates-industry/.

Herzog, Alexander, and Slava J. Mikhaylov. 2017. "Database of Parliamentary Speeches in Ireland, 1919–2013." *2017 International Conference on the Frontiers and Advances in Data Science (FADS)*, 29–34. https://doi.org/10.1109/FADS.2017.8253189.

Hestness, Joel, Sharan Narang, Newsha Ardalani, Gregory Diamos, Heewoo Jun, Hassan Kianinejad, Md. Mostofa Ali Patwary, Yang Yang, and Yanqi Zhou. 2017. "Deep Learning Scaling Is Predictable, Empirically." Preprint, submitted December 1. http://arxiv.org/abs/1712.00409.

Howard, Christian. 2018. "Introducing Google AI." Google AI Blog, May 7. http://ai.googleblog.com/2018/05/introducing-google-ai.html.

Huang, Dan. 2018. "How Much Did AlphaGo Zero Cost?" *Dansplaining* (blog), March. https://www.yuzeh.com/data/agz-cost.html.

Jovanovic, Boyan, and Peter L. Rousseau. 2005. "General Purpose Technologies." Working Paper. Cambridge, MA: National Bureau of Economic Research.

Kaggle. 2019. "State of Data Science and Machine Learning 2019." Kaggle. https://www.kaggle.com/kaggle-survey-2019.

Kaiser, Caleb. 2020. "Too Big to Deploy: How GPT-2 Is Breaking Production." *Medium*, February 2. https://explainjay.wordpress.com/2020/02/02/too-big-to-deploy-how-gpt-2-is-breaking-production/.

Kohlberger, Timo, and Yuan Liu. 2020. "Generating Diverse Synthetic Medical Image Data for Training Machine Learning Models." Google AI Blog, February 19. http://ai.googleblog.com/2020/02/generating-diverse-synthetic-medical.html.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey Hinton. 2012. "ImageNet Classification with Deep Convolutional Neural Networks." In *Proceedings of the 25th International Conference on Neural Information Processing Systems* 1: 1097–1105. Red Hook, NY: Curran Associates Inc.

Kumar, Ashutosh, Arijit Biswas, and Subhajit Sanyal. 2018. "eCommerceGAN: A Generative Adversarial Network for E-Commerce." Preprint, submitted January 10. http://arxiv.org/abs/1801.03244.

Kuznetsova, Alina, Hassan Rom, Neil Alldrin, Jasper Uijlings, Ivan Krasin, Jordi Pont-Tuset, Shahab Kamali, Stefan Popov, Matteo Malloci, Alexander Kolesnikov, Tom Duerig, and Vittorio Ferrari. 2020. "The Open Images Dataset V4: Unified Image Classification, Object Detection, and Visual Relationship Detection at Scale." *International Journal of Computer Vision* 128: 1956–1981. https://doi.org/10.1007/s11263-020-01316-z.

Laanait, Nouamane, Joshua Romero, Junqi Yin, M. Todd Young, Sean Treichler, Vitalii Starchenko, Albina Borisevich, Alex Sergeev, and Michael Matheson. 2019. "Exascale Deep Learning for Scientific Inverse Problems." Preprint, submitted September 24. http://arxiv.org/abs/1909.11150.

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

Lee, Kai-Fu. 2018. *AI Superpowers: China, Silicon Valley, and the New World Order*. Boston: Houghton Mifflin Harcourt.

Levy, Steven. 2017. "Bill Joy Finds the Jesus Battery." *Wired*, August 16. https://www.wired.com/story/bill-joy-finds-the-jesus-battery/.

Li, Xuan, Kunfeng Wang, Yonglin Tian, Lan Yan, and Fei-Yue Wang. 2017. "The ParallelEye Dataset: Constructing Large-Scale Artificial Scenes for Traffic Vision Research." Preprint, submitted December 22. http://arxiv.org/abs/1712.08394.

Mahidhar, Vikram, and Thomas H. Davenport. 2018. "Why Companies That Wait to Adopt AI May Never Catch Up." *Harvard Business Review*, December 6. https://hbr.org/2018/12/why-companies-that-wait-to-adopt-ai-may-never-catch-up.

Menant, Sebastien, and Pracheer Gupta. 2017. "Bring Machine Learning to iOS Apps Using Apache MXNet and Apple Core ML." *Amazon Web Services* (blog), September 6. https://aws.amazon.com/blogs/machine-learning/bring-machine-learning-to-ios-apps-using-apache-mxnet-and-apple-core-ml/.

Metz, Cade. 2019. "A.I. Is Learning from Humans. Many Humans." *New York Times*, August 16. https://www.nytimes.com/2019/08/16/technology/ai-humans.html.

Mucha, Tomasz, and Timo Seppälä. 2020. "Artificial Intelligence Platforms: A New Research Agenda for Digital Platform Economy." ETLA Working Paper no. 76. Helsinki: Elinkeinoelämän Tutkimuslaitos. https://www.etla.fi/wp-content/uploads/ETLA-Working-Papers-76.pdf.

Murgia, Madhumita. 2019a. "AI Academics under Pressure to Do Commercial Research." *Financial Times*, March 13.

Murgia, Madhumita. 2019b. "AI's New Workforce: The Data-Labelling Industry Spreads Globally." *Financial Times*, July 24. https://www.ft.com/content/56dde36c-aa40-11e9-984c-fac8325aaa04.

Nadella, Satya. 2018. "Embracing Our Future: Intelligent Cloud and Intelligent Edge." Microsoft News Center, March 29. https://news.microsoft.com/2018/03/29/satya-nadella-email-to-employees-embracing-our-future-intelligent-cloud-and-intelligent-edge/.

Nahles, Andrea. 2018. "Die Tech-Riesen des Silicon Valleys gefährden den fairen Wettbewerb" [The Tech Giants of Silicon Valley Endanger Fair Competition]. *Handelsblatt*, August 13. https://www.handelsblatt.com/meinung/gastbeitraege/gastkommentar-die-tech-riesen-des-silicon-valleys-gefaehrden-den-fairen-wettbewerb/22900656.html.

Niklas, Jędrzej, and Lina Dencik. 2020. "European Artificial Intelligence Policy: Mapping the Institutional Landscape." Working Paper. Cardiff: Data Justice Lab. https://datajusticeproject.net/wp-content/uploads/sites/30/2020/07/WP_AI-Policy-in-Europe.pdf.

Nisselson, Evan. 2018. "Deep Learning with Synthetic Data Will Democratize the Tech Industry." *TechCrunch*, May 11. http://social.techcrunch.com/2018/05/11/deep-learning-with-synthetic-data-will-democratize-the-tech-industry/.

Peng, Tony. 2019. "Data Annotation: The Billion Dollar Business behind AI Breakthroughs." *Synced*, August 28. https://medium.com/syncedreview/data-annotation-the-billion-dollar-business-behind-ai-breakthroughs-d929b0a50d23.

Pleatsikas, Christopher, and David Teece. 2001. "The Analysis of Market Definition and Market Power in the Context of Rapid Innovation." *International Journal of Industrial Organization* 19 (5): 665–693.

Rao, Anand S., and Gerard Verweij. 2017. "Sizing the Prize: What's the Real Value of AI for Your Business and How Can You Capitalise?" New York: PwC. https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf.

Royal Society and British Academy. 2018. *The Impact of Artificial Intelligence on Work*. London: Royal Society and British Academy. https://royalsociety.org/-/media/policy/projects/ai-and-work/evidence-synthesis-the-impact-of-AI-on-work.PDF.

Simonite, Tom. 2018. "Some Startups Use Fake Data to Train AI." *Wired*, April 25. https://www.wired.com/story/some-startups-use-fake-data-to-train-ai/.

Srnicek, Nick. 2016. *Platform Capitalism*. Cambridge: Polity Press.

Sun, Chen, Abhinav Shrivastava, Saurabh Singh, and Abhinav Gupta. 2017. "Revisiting Unreasonable Effectiveness of Data in Deep Learning Era." *2017 IEEE International Conference on Computer Vision (ICCV)*, 843–852. https://doi.org/10.1109/ICCV.2017.97.

Sutton, Richard. 2019. "The Bitter Lesson." *Incomplete Ideas* (blog), March 13. http://www.incompleteideas.net/IncIdeas/BitterLesson.html.

Toole, Jameson. 2019. "Deep Learning Has a Size Problem." *Heartbeat*, November 5. https://heartbeat.fritz.ai/deep-learning-has-a-size-problem-ea601304cd8.

Trajtenberg, Manuel. 2018. "AI as the Next GPT: A Political-Economy Perspective." Working Paper 24245. Cambridge, MA: National Bureau of Economic Research.

Varian, Hal. 2018. "Artificial Intelligence, Economics, and Industrial Organization." Working Paper 24839. Cambridge, MA: National Bureau of Economic Research.

Virno, Paolo. 2004. *A Grammar of the Multitude*. Los Angeles: Semiotext(e).

Weber, Steven. 2017. "Data, Development, and Growth." *Business and Politics* 19 (3): 397–423.

Published by the MIT Press.

A copublication with
International Development Research Centre
PO Box 8500
Ottawa, ON K1G 3H9
Canada
www.idrc.ca/info@idrc.ca

The research presented in this publication was carried out with the financial assistance of Canada's International Development Research Centre. The views expressed herein do not necessarily represent those of IDRC or its Board of Governors.

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Graham, Mark, 1980– editor. | Ferrari, Fabian, editor.
Title: Digital work in the planetary market / edited by Mark Graham and Fabian Ferrari.
Description: Cambridge, Massachusetts : The MIT Press, 2022. | Series: The MIT Press-International Development Research Centre series | Includes bibliographical references and index.
Identifiers: LCCN 2021037262 | ISBN 9780262543767 (paperback)
Subjects: LCSH: Employees—Effect of technological innovations on—Case studies. | Industrial productivity—Effect of technological innovations on—Case studies. | Electronic commerce—Case studies.
Classification: LCC HD6331 .D527 2022 | DDC 331.25—dc23/eng/20211208
LC record available at https://lccn.loc.gov/2021037262