Berkeley
UNIVERSITY OF CALIFORNIA

BERKELEY ROUNDTABLE ON
THE INTERNATIONAL ECONOMY

BRIE Working Paper
2021-1

Algorithms, Data, and Platforms: The Diverse Challenges of Governing AI

Mark Nitzberg and John Zysman

Funding: This research was funded in part by the Ewing Marion Kauffman Foundation with additional support from the German Ministry of Labour (BMAS).

Acknowledgements: We appreciate and wish to emphasize the substantial contribution of Camille Carlton.

Algorithms, Data, and Platforms:
The Diverse Challenges of Governing AI

Mark Nitzberg
Executive Director,
Center for Human-Compatible Artificial Intelligence (CHAI);
Director of Technology Research,
Berkeley Roundtable on the International Economy (BRIE);
Head of Strategic Outreach,
Berkeley AI Research (BAIR)
University of California, Berkeley
nitz@berkeley.edu

&

John Zysman^
Professor Emeritus,
Department of Political Science;
Co-Founder and Co-Director,
Berkeley Roundtable on the International Economy (BRIE)
University of California, Berkeley
zysman@berkeley.edu

^Corresponding Author

This paper has been invited to the Journal of European Public Policy as part of the Special Issue, 'Governance of Artificial Intelligence.'

Abstract Artificial Intelligence (AI) poses interwoven challenges. Defined as technology that uses advanced computation to perform at human cognitive capacity in some task area, AI must be regulated in the context of its broader toolbox - algorithms, data and platforms - and its regulation must be sector-specific. Establishing national and community priorities on how to reap AI's benefits, while managing its social and economic risks, is an evolving debate. Digital Platform Firms are a fundamental driver of AI tools: they dominate the playing field and often pursue priorities outside the frames of the public sector and of civil society. While its governance is critical to national success, AI pries open a Pandora's box of questions that sweep across the economy and society, engaging diverse communities. Rather than a single, global ethical framework, one must consider how to pursue objectives of interoperability amongst nations with quite different political economies.

Keywords: Artificial Intelligence, Governance, Emerging Technologies, Digital Platforms

Introduction

Imaginations often run wild with the vast possibilities of artificial intelligence (AI). Many believe that AI is already capable of most "thinking human" tasks (Marr, 2019), while others see a jobless dystopia (Zysman et al., 2019). The reality is that today's AI systems are part of an ever evolving "technology toolbox" that, for now, are only effective in well-specified problem domains. Our working definition of AI is *technology that uses advanced computation to perform at human cognitive capacity in some task area.* This has critical implications for AI governance,[1] as policy- and decision-makers must focus on the distinctive and powerful problems posed by AI's *current* capabilities, as opposed to the theoretical prospect of "General AI", that extends AI from *some* to *all* task areas. Nonetheless, governance in a digital era is about the challenges represented not just by AI, but by the technology toolbox as a whole.

Governing new general-purpose technologies, such as steam, electricity, and now AI, and establishing the rules and norms for their development, is difficult as their possibilities, applications, and risks only emerge in their early years. Whom these technologies benefit, whom they harm, and how they do so is rarely immediately evident, and often underestimated and exaggerated. The rapid evolution of AI's underlying technologies pressures public authorities to keep up, both in their understanding and in their rule making. One might, as a consequence, suggest that the ideal path forward would involve careful, iterative steps addressing issues as they become clearer and drafting regulation that is actionable yet mouldable. However, as the history around technology rule making has taught, the early decisions regarding one aspect of technology will constrain choices on other aspects.[2] The early decisions create distinctive paths of development, technology trajectories. As Paul Allen David (1986) so wisely put it, we are blind giants pushing through these narrow windows. There is a narrow window for rulemaking, before decisions of investment and deployment create enduring trajectories of technological and political economy development.

Governing AI poses a set of interwoven challenges. These challenges begin with agreeing on what AI actually is. In an earlier piece,[3] Zysman and Nitzberg (2020) argued:

> "The core story about governance in a digital era is not about the rules of particular tools, but the challenges represented by a new toolbox. That toolbox includes big

[1] Let us not scuffle about a precise definition of "governing" or governance, since such precision is not needed for this discussion. Something like this will do for now:

> *Governance is the way rules, norms and actions are structured, sustained, regulated and held accountable. The degree of formality depends on the internal rules of a given organization and, externally, with its business partners. As such, governance may take many forms, driven by many different motivations and with many different results. For instance, a government may operate as a democracy where citizens vote on who should govern and the public good is the goal, while a non-profit organization or a corporation may be governed by a small board of directors and pursue more specific aims. (Wikipedia, accessed February 14, 2021)*

[2] For example, a suit under the Sherman Antitrust act of 1890 broke up Standard Oil in 1911. This kept gasoline prices low. At the same time, the regulatory setting enabled the introduction of Ford's assembly line. Together these literally fueled Ford's rise, selling their Model T for under $700, to become the largest and most profitable manufacturer in the world by 1912, and incidentally closing off the path to electric cars as an alternative.

[3] The earlier piece was *Governing AI: Understanding the Limits, Possibilities, and Risks of AI in an Era of Intelligent Tools and Systems*, published by the Woodrow Wilson Center. This paper builds on this earlier publication, expanding the debate beyond the questions and issues of governance to implementing governance itself.

> data, powerful computing capacity, algorithms, and software in general. AI, in its several current manifestations, is one tool amongst many, a truly powerful tool with an ever-expanding set of applications. … The crucial governance questions are about how we deploy the tools—to what ends? with what benefits, and to whom? with what risks?" (pp. 2-3)

There is enormous variation evident in national and community priorities and hence choices about AI. The differences are not just on how to resolve these challenges, but on the preferred outcomes of governance. Part of that struggle is whether the choices will be public or private. We must recognize that AI deployment by Digital Platform Firms (DPFs) is resettling contemporary markets as well as the uses of data and the algorithms that underpin the AI challenges. The remarkable power of DPFs warrants a critical look at who owns, and thus controls, the application of AI in so many domains of the suite of AI technologies.

We develop our argument as follows: *First*, we ask, what is AI? What is it capable of? And what are its limits? We recognize, *second*, that AI does not stand alone; in fact, it is part of a toolkit which includes data, platforms, and computing. *Third*, there is a tension between the search for general rules and the particularities of regulation in specific contexts. When approaching these technologies, it is critical to consider that their application raises issues and concerns specific to particular economic and social domains. Those sectoral particularities ought to greatly influence governance. Therefore, those who would govern AI must ask whether to seek overarching rules or domain specific regulation. They must consider which issues fail to be addressed effectively by tool-based or domain-based rules. Importantly, *fourth*, there are varying interests across public, private, and civil sectors, with DPFs having overwhelming influence through enormous investment and deployment at scale of the technological elements that comprise the AI toolkit. What does this power structure mean for governance? Finally, *fifth*, balancing competing priorities, balancing innovation and risk, in diverse jurisdictions will be difficult. With real concerns that the digital economy will fracture into two, three, or four distinct systems, should the objective in international discussion be a single marketplace with relatively similar rules, or should we settle for a form of interoperability?

AI has great potential but involves significant risk. The question for governance is how to encourage the potential while minimizing the risk. So, let us get started. What is AI?

Part I: What is AI and What is it Capable of?

Despite Artificial Intelligence (AI) becoming increasingly utilized within industry, studied within academia, and of growing concern by public institutions, it lacks a standard definition. The definitions of AI vary as greatly as society's perceptions of it, considering a range of abilities from "information processing" to "decision making"; however, most definitions refer to its "human intelligence."[4] The definition we use throughout this paper of AI is *technology that uses advanced computation to perform at human cognitive capacity in some task area*. This definition has two critical aspects that we explore further: "advanced computation" and "in some task area."

[4] See, for example, McCarthy et al., 1955; Minsky, 1968; Gardner, 1999; McCarthy, 2007; Russel and Norvig, 2010; HLEG, 2019.

Machine learning, the principal computer science tool enabling contemporary AI is, as Zysman and Nitzberg previously stated, "fundamentally a form of context-dependent statistical inference, and such has its limits" (Zysman and Nitzberg 2020, p. b). The AI milestones we've seen today — audio speech transcription, image recognition, intent inference from search terms in search engines, behaviour pattern clustering, content selection leading, e.g., to addictive video-feeds, protein folding, etc. — have been enabled by data-driven machine learning (ML). Still, data-driven machine learning is capable of much more and is expected to continue to automate many more cognitive tasks and physical tasks in the coming years.

First, we focus on the initial part of our definition, advanced computation, especially the data-driven machine learning techniques that have enabled the "AI breakthroughs" of the last 10 years. Then, we dive into the qualifier of our definition, "in some task area," which highlights the limits of contemporary AI.

Understanding AI: From the Technical Underpinnings to Its Rise

Machine learning, the principal computer science tool enabling contemporary AI, is the study of computer algorithms which, utilizing vast amounts of data for training, make predictions, determine actions, and improve automatically through experience. Whereas ordinary algorithms are designed to produce a particular output response to given inputs, ML algorithms "learn" to tailor their outputs for any given input by iteratively adjusting many—often millions—of parameters based on large sets—often billions—of inputs. Current ML techniques include supervised learning[5] which imitates past observations; reinforcement learning[6] which maximizes a goal through trial and error; and unsupervised learning which finds patterns, from grouping like with like ("people who liked Moby Dick also liked …"). Recent developments include novel "deep network" architectures that enable "self-supervised learning," which have enabled the astonishingly sophisticated behaviour GPT3.[7] However, as the techniques underlying contemporary AI systems are based simply on input-output responses, AI as we know it is limited to "Narrow AI," or AI which can accomplish well-defined tasks.[8]

[5] From vast corpora of data marked up by humans with features, identities, or labels, a system iteratively improves ("learns") a large set of probabilities in order to predict the most likely output labels for new, never-before-encountered input data. The class of techniques, called "deep learning," combining these data structures and training techniques, enabled a leap in image recognition around 2011 (see e.g., Krizhevsky et al., 2012) that put AI back on the map – some 40 years after these computational models had first been proposed (McCulloch and Pitts, 1943).

[6] A system uses large scale random trial-and-error to maximize a score, such as board-position strength in chess, wind turbine efficiency, stock portfolio value, or robot motion effectiveness. When combined with deep networks, so-called deep reinforcement learning can yield systems like AlphaZero that learn, through trial-and-error, to play games like Go beyond human capacity.

[7] GPT-3 is a key example of a new type of ML architecture that enables systems to "learn" features and structures of a class of data – in this case, written language – without human-labeling. The architecture enables so-called "self-supervised learning," in which written texts are mined over and over by covering and revealing certain words, to train a language model to predict those covered words. The probabilities of grammar, usage, and style emerge in the system after it is exposed to vast corpora of books, online news sources, and publications.

[8] While this paper focuses on Narrow AI, it does not dismiss the eventual development of General AI (AGI). In fact, there is considerable effort being made to develop and assure the safety of AGI. See Nitzberg's own organization,

Made possible by ML, today's AI systems have a wide range of capabilities with both vast and critical applications. Within these narrowly defined tasks, AI systems achieve a speed and precision unmatched by any human. For example, AI systems can recognize objects in a picture, transcribe speech, diagnose illnesses from symptoms and test results, maximize content engagement, and reproduce artistic styles just as fast and accurately, if not more, than most humans (Cohen, 2016; Deutsch, 2016; Eckersley & Nasser, 2017; Joshi et al., 2017). In domains which require the analysis of extremely large data sets to discover and reproduce patterns, AI is undoubtedly unmatched. Yet, AI does not stand alone.

AI functions as part of a "technology toolbox" made up of algorithms, data, and platforms. These tools do not simply exist alongside each other; instead, they are an interconnected system with feedback loops, reinforcing the behaviour, and outcomes, they create. To illustrate, most see algorithms as the heart of AI; however, these algorithms utilize, and sometimes learn, according to the data that they are fed. One might argue that the largest "data collectors" to date are the large digital platform firms (DPFs)[9] which, in many cases, go on to implement their own algorithms. Conceptualizing AI as part of this technology toolbox is critical not just for understanding the technical aspects, but for governance.

Over the last two decades, the astonishing breakthroughs in AI, and the systems and applications that deploy the techniques, have been driven by the dramatic expansion of various technology infrastructure, such as widely available and affordable internet and computer processing power. The mass data sets necessary for training ML algorithms, the platforms which gather such data sets, and the algorithms which are enabled by data and, many times, developed by said platforms, have all been catalysed by the rise in internet and computer processing power. This technology infrastructure has enabled AI to operate at unprecedented levels, as interactive connected devices reach billion-user and trillion-transaction scale. As of the end of 2020, each minute, Google processed 270 million searches; Facebook recorded 1 million logins; Twitter posted 511,200 tweets; and YouTubers streamed 4,500,000 videos — all which, in total, contribute to the 2.5 quintillion bytes of the data believed to be generated daily (Karki, 2020). The generation of data has risen exponentially with the rise of available internet; over 90% of the total global data collected has taken place since 2016 alone (Marr, 2018; Karki, 2020).

AI powered operations, such as information (e.g., weather, GPS location, health metrics, and purchasing preferences), communication, and influence, have reached a scale beyond that of prior eras, that our norms and the laws on which they were based, simply do not apply in an obvious way. Never before could one person pass a precision-tailored message to a billion others in the course of a day, nor every square centimetre photographed on every paved street, available to anyone anywhere else on the face of the planet. The implications of this scale for governance

Center for Human Compatible Artificial Intelligence (CHAI), as well as Open AI, Deepmind, Google Brain, Facebook A.I. Research, AGI Innovations, Microsoft Research Lab, Apprente, and Kimera Systems.

[9] A research emphasis of the Berkeley Roundtable on the International Economy focuses on the diverse ways in which DPFs are rewiring the economy and society. See Kenney and Zysman (2016), Bearson et al. (2020), Kenney and Cutolo (2020), and Kenney and Zysman (2019). Kenney et al. (forthcoming) illustrate how DPFs, once planted in traditional economic sectors, follow multiple simultaneous expansion paths to increase their presence and power across new industries.

are significant — not only do we need new interpretations of existing laws, we also need new laws which better reflect the limits, potentials, and risks of contemporary AI.

The Limits of Contemporary AI

Despite AI's impressive achievements and often anthropomorphic portrayal, these systems fall short of human intelligence in many respects (Marcus, 2020). For our purposes AI lacks three fundamental aspects of human cognition and interpretation: context, narrative, and worldview. By *context,* we mean that faculty of human intelligence to broaden one's view of a task at hand to situate it among our models of how the world works, whether the setting is familiar or completely new. Contemporary AI's "contextualization" is confined to the scope of the system's inputs. Chess software is only aware of the board, pieces, moves — all things chess. It is not aware of the flagging energy of the opponent, for example, something which a human player is able to factor in. A news recommender system's context is limited to a user's "preferences" based on past behaviour, in the weakly broader context of advertiser and platform objectives, leaving out the broader context of what is beneficial to users, society, or the economy (Marcus and Davis, 2019). This higher-level awareness of context underlies all human endeavours regardless of familiarity. We understand *narrative* as constituents of a greater *worldview,* only possible when you can apply context and bring judgement — the ability to step back from the immediate domain and understand something's significance to the broader world — to bear in complex situations.[10] For example, narratives such as "we should have a four-day work week" or "we should only allow for sustainable fishing" are stitched together to create our larger worldviews. However, without a higher-level awareness of context, AI systems simply cannot form narratives or a worldview.

An AI powered vehicle can deduce from visual and other inputs that a ball is rolling into the street; however, being limited to correlations and physics, having no concept of the use or purpose of a ball – other cars, trees and balls are simply obstacles - the AI is unable to deduce that the ball may be followed by child at play, the "cause" of the ball's trajectory (Knight 2020). Lacking this context, the autonomous vehicle would fail to stop in anticipation of a child running into the street. Finally, AI systems cannot construct their own narratives and worldview. If an accident were to occur, the autonomous vehicle would have no means to comprehend the trauma inflicted on both the victims and the passengers, much less to "show remorse." Together, the lack of context, narrative, and worldview pose obvious concerns for the deployment of AI systems, particularly those built to operate in a physical environment.

These limits, which differentiate AI from human intelligence, both underline its potential and usher in unavoidable risks of use. So, what exactly are the potential and risks of contemporary AI? As AI systems are unmatched in their ability to analyse large data sets for the completion of set tasks, the possibilities of their application lie in the previously unimaginable scale and precision they can achieve. However, it is precisely this scale, combined with contemporary AI's foundational model that optimizes for fixed, well-defined objectives, which enable both unintended consequences (e.g., misaligned algorithmic goals aiming to increase engagement actually increasing extremism) and malicious uses (e.g., autonomous weapons, precision propaganda, and increasingly sophisticated cyberattacks) (Brundage et al., 2018). Given

[10] George Akerlof and Robert Shiller's extensive writing emphasize the importance of narrative. Two books capture their positions: Shiller (2019) and earlier Akerlof and Shiller (2009).

the staggering breadth of potential applications of AI, the possibilities and risks are best understood within specific contexts. We do just that, exploring the possibilities and risks of AI within the context of governance, in the following sections.

Part II: Governing AI: The Issues

AI is a tool with powerful but limited capabilities. We revisit the governance question: *how do we most effectively maximize the benefits of AI while avoiding the harms?* The ideal approach to governing AI and the related technology toolbox would be iterative and contextually-informed; however, this is not feasible as we have a narrow window for implementing technology governance.

We provide overarching considerations on how to approach governance, asking what should be governed, who is (and should be) involved, and what we can expect moving forward. The economic and social domains in which AI operates are the context in which rules and norms of governance will be built. To consider how to apply AI in these contexts practically, we must approach governance through sector-specific rules. Still, there are concerns that cut across domains which require addressing the toolbox as a whole. In addressing the toolbox, decisions about DPFs powerfully influence the deployment of AI now, and will shape the debate going forward. At issue, is whether public or private interests regulate AI. As it stands today, somewhat overstated, but not entirely incorrect, regulation is, in many domains, de facto determined by these powerful DPFs. Since AI applications reach across borders, the issue of governance is inherently international. Communities differ in their goals and priorities. For example, in Germany, property owners have a right to exclude their building façades from Google's globally accessible "Street View" system, whereas Americans have no such right, at least for now. Consequently, a single global deal, international governance of AI, seems unlikely, although, there may be some particular domains where real agreement is possible. The overriding question will be some system of interoperability so that tools and applications can work across boundaries. In our Street View example, Google allows German owners to opt out, enabling compliance with multiple sets of regulations.

A Focus on Outcomes: AI in the Economy and Community

What benefits, or outcomes, of AI's deployment do we desire? What type of society do we want to live in? How does this new technology toolbox influence these goals? We consider two arenas, economy and community. While we cannot elaborate each area in depth, we strive to illustrate the character and range of challenges that the AI toolbox presents.[11]

Economy

[11] For our purposes and discussion, we set aside military and geo-strategic questions as well as the economic debates around promoting innovation.

For governance, the deployment of AI obviously raises questions about the impact on growth and labour. While noting that the race for AI superiority is certainly driven by the strategic goals of the major powers such as China, US, Russia, and European countries, we set aside concerns over cyber security and autonomous weapons for a separate discussion.

Capturing the benefits of AI technologies will require the successful deployment of the full suite of intelligent tools. While AI technological leadership[12] is arguably essential for strategic goals, its role in economic development and productivity growth is debated. Technical advances quickly diffuse across borders. Leading-edge AI research continues the scientific tradition of global dissemination through peer-reviewed journals and conferences, while the DPFs reduce new developments to practice and use a degree of trade secrecy to compete.

The key to capturing the economic possibilities of a new general purpose technology (GPT) is its successful deployment, not the simple investment in advancing cutting edge applications. This may seem obvious, but much of policy debate is about AI leadership that in effect focuses on cutting edge applications and not deployment. The introduction of new GPTs have historically seen radical economic restructuring, in some views driving growth surges, in others generating the avenues for sustained investment. Critically, this restructuring transcends the initial technology.[13] It involves investment and innovation in the suites of complementary technology (in this case broadband and WIFI for communication); significant reorganization of the workforce beyond task and role substitutions (Manyika et al., 2017); and investment in human capital for skill-complementing and upgrading.

Importantly, AI is, quite evidently, distinct from earlier GPT in one important dimension: AI is the automation of certain tasks of human cognitive capacity. Factories have, over the past decades, been transformed by digital tools from NC machinery through robotics. While AI may refine the application of such tools in traditional manufacturing, and indeed in agriculture and services, it will most powerfully impact the service sector. This is not just applications in restaurants or in management of cleaning services. Most importantly, there is the creation of entirely new kinds of information services as well as the conversion of value creation from goods industries into ICT enabled services (Nielsen et al., 2013; Zysman et al., 2013). Search engines, retail sales, and the sale of tires or engines as "services" are just part of the list. For instance, AI has already managed to eliminate the need for CPAs for a certain level of tax advice (Kravitz, 2020). Now, it's being utilized within the legal field to complete due diligence, research, and bill hours; within the hospitality industry to prepare and deliver foods; and across various sectors to help manage customer service requests. That said, AI embedded in platforms permits Amazon to shift retail from storefront to online and with that transform clerk jobs into warehouse jobs. The consequences are only beginning to unfold, but at least for now – with variation across countries –middle skilled employment seems to be eroding (OECD, 2020; Manyika et al., 2020). That though is a focus for another day.

[12] Here, we define AI technological leadership as distinct from access to leading edge research.
[13] This has been the case historically. For instance, the technical innovation of the railroad was dependent on complementary technologies (e.g., metal for rail and engines), foundational technologies (e.g., steam), and operational training.

Policy choices, governance, will shape the distributional consequences of this new technology toolbox, not just its pace of adoption. A diverse array of choices, not all directly linked to work tasks, will matter. Consider *infrastructure*. Will the broadband infrastructure be in place in rural communities, or available at viable costs for less privileged communities? The consequences for implementing AI tools and services on farms, on the education of the community, will be significant. Consider *labour markets*. In the US, the debate is whether so-called gig workers will be treated as contractors out there on their own, employees with employment rights, or some third category. More generally in the advanced countries, the questions will be labour rights, claims on welfare and training. Consider *tax policy.* Will there be tax breaks for capital investment that encourage substitution of AI driven equipment for labour, the classic substitution of capital for labour? Alternatively, will workforce training be a tax deduction that encourages complementary division of tasks between people and machines? Indeed, consider *research policy.* Do we consciously invest in technology and deployment strategies, and management education, that upskills work, seeking the most complementary divisions between people and machines?

Government policy and the choices of firms and public sector agencies that deploy the technology, not the mere existence of AI, will determine the outcomes we see in each of these debates.[14]

Community

The effects of AI on our communities will depend, likewise, on choices about the specific applications of AI technologies, making the consideration of community objectives and norms critical. AI is being applied in staggeringly diverse sectors. The breadth of applications tempts us to define overarching general rules. Yet the particular issues raised are quite often specific to the sector or domain. Evidently, there will be both overarching debates about data or facial recognition across sectors, and domain-specific debates about the use of patient data and which firms should have access to retail search data. It is common knowledge that AI is used in healthcare to develop new drugs and recommend treatments; in criminal justice to determine sentencing; and in our news feeds and browsers, shaping both our online and offline social discourse and consumer preferences. What should the rules and norms be?

Values, sometimes implicit, will be formalized as program code, not just law, in the development of AI applications. There are battles about which norms and values to embed in law and code, and when to apply which rules, an argument that Larry Lessig classically clarified and which others have built on (Lessig, 1999).[15] We object to general facial recognition for surveillance, but law enforcement has found it an effective tool to identify and arrest the perpetrators of the January 6 attack on the U.S. Congress. That is only the beginning. Even if a community could agree, today, on values across an array of domains, community norms will evolve. Today's norms reflect a snapshot of social practice put into regulation *at a given moment* in time. As social practice evolves, community preferences will change, as should the political

[14] This has been thoroughly argued by Zysman and colleagues. See Zysman, Kenney, and Tyson, 2019.
[15] Discussed in a range of works including Fleischmann and Wallace, 2006; Zunger, 2016; Groth and Nitzberg, 2018; and Christian, 2020.

regulation to reflect this. Take, for instance, the recent powerful awakening, and corresponding shift in social norms, brought about by the Black Lives Matter (BLM) movement. It has evolved from a social media hashtag (following the acquittal of Travon Martin's murder) to radical justice call inspiring movements in over sixty countries, which was recently nominated for the Nobel Peace Prize (Black Lives Matter, 2021). It is important then to recognize the continuing evolution of social values, and the risk not only of entrenching todays' values in law, visible perhaps, but less visibly in code.

There is another risk which requires addressing, even if it is not elaborated. In the pursuit of AI enabled efficiency, we may confuse ourselves about the objectives we want to pursue. As a general matter, we must not allow AI to entrench objectives and obscure the political debate about values. Smart cities, we are told often, will improve urban living and governance, but that begs the question of smart toward what ends. Perhaps AI can help optimize governance, such as efficient trash collection or traffic flows. But, with its limitations (i.e., AI's lack of context, narrative, and worldview), it cannot answer the questions of whose garbage should be collected or through whose neighbourhood traffic is directed.

Agreeing on when, where, and how we'd like AI to be used across its vast applications, while also considering the potential consequences of decisions which reflect temporary social norms, is no trivial feat. How then, do we design a governance strategy that addresses the multitude of AI uses while being flexible enough to adapt to our evolving norms?

Part III: Governing AI: In Action

Our continuing mantra is that the values which should guide governance, differ across the public and civil sectors, and within the private sector by domain. That complicates the path forward. Today, powerful digital platform firms are the ad hoc, but pervasive, governors of AI technology and its applications. The decisions which we make regarding the toolbox, in particular about platforms, will determine whether public or private interests regulate AI moving forward.

A Two-Pronged Approach to AI Governance

Evidently from our discussion, there are two approaches to AI governance, focus on specific applications or generate overarching principles and rules. Certainly, specific cases can suggest general rules and general concerns and rules can inform, and shape, the choices we make in particular situations.

The same AI application in different contexts will generate gains and consequences distinct both to the domain and purpose in which it is applied — dictating the need to focus primarily on sector-specific applications. The use of AI may be less problematic in certain sectors than others. Take, for example, digital advertisements. A targeted retail ad on fuzzy

slippers which draws an individual into a cycle of snow boot advertisements, is of less significance than a targeted political ad leading an individual into a cycle of content claiming the basement of a pizza parlour is host to a satanic child-trafficking ring.[16] In both scenarios, algorithms utilize vast data points to personalize recommended content within social platforms; yet, the outcome in the retail sector, as opposed to the political one, is far less concerning.

Nonetheless, there will be concerns that cut across many social domains and economic sectors. In focusing solely on sector-specific applications, there is the potential that issues which reappear across sectors may be hidden or underestimated because the issue is not central in a particular domain. Issues, such as bias in data and algorithmic transparency, may be most evident and easily identified in some domains, but are important throughout society. Choices about data privacy and surveillance arise throughout. As we noted previously, considering the entire toolbox means focusing on algorithms, data, and platforms as well as the technology infrastructure -- the internet and computing power -- which grants power to these tools.

The challenges of AI governance require policy makers to have a significant understanding of both AI's technical aspects (e.g., its strengths, limits, and its role in the toolbox) as well as the social and economic contexts in which it is situated. Indeed, policy makers must consider that despite the objective of contemporary AI systems being narrow (e.g., score an applicant, increase engagement, etc.) the scale at which these systems are deployed will compound the impact of such systems, spreading risks far beyond initial domains. This expertise may be difficult to replicate across multiple sectors, but will be necessary to address the complex challenges of governance. We explore two cases, bias and disinformation, which demonstrate such challenges, below.

Bias

The application of algorithmic decision making has resulted in unintended bias, particularly against underrepresented communities, such as people of colour, indigenous people, LGBTQ+ individuals, people with disabilities, and women. Bias in AI can stem from the design and deployment of an algorithm. Certainly, algorithms designed by Alexandria Ocasio-Cortez (AOC), the progressive democratic congresswoman from New York City, and Jeff Session, Donald Trump's attorney general, would differ dramatically. Bias results from the data that is fed into the analysis. In general, systems trained on data that captures past decisions and behaviour, necessarily capture the biases of that past era. AI can also amplify existing biases, as algorithmic decision making systems, which capture existing biases from inputs to inform outputs, are deployed at an exponential scale.

Bias in algorithmic decision making spans all sectors. Health-prediction, credit application, bail determination, and hiring algorithms, have all been found to demonstrate significant levels of racial bias, further perpetuating social and economic inequalities at scale

[16] "PizzaGate" is a widely debunked conspiracy theory, a prominent example of the politically motivated disinformation spread prior to the 2016 U.S. presidential election. (Aisch et al., 2016).

(Agwin, 2016; Dastin, 2018; Klein, 2020; Kaushal, 2020). In a recent example, a widely used healthcare algorithm, aimed at determining "high-risk" patients in need of additional care, was found to exhibit such high levels of bias that it reduced the necessary number of Black patients identified for "high risk care" by over half (Obermeyer, 2019). The algorithm utilized previous healthcare spending data as a proxy for medical need. This data failed to capture a systemic health inequity — the fact that black patients spend more on active health interventions than white patients — meaning that similar spending did not portray similar levels of health (Ibid.). Thus, the algorithm, intended to efficiently provide care recommendations, instead furthered health inequalities at scale.

To generalize, Cathy O'Neil (2016), points to the technological design of scalable algorithms and historical data creating an unchallengeable set of conclusions. If algorithms continue to be deployed without any consideration for the biases present in their designs and training data, they will continue to amplify existing inequalities. Often, AI is simply obscuring the source of bias instead of uncovering it. Thus, AI may become a tool which encourages a dystopia, instead of being harnessed to achieve our desired societies. Yet, some have pointed to the potential for AI to reduce, or at least make bias more evident. For instance, the video functionality of smartphones has recently been used to document racism and violence within policing. Perhaps, as applications are deployed, AI tools, with human guidance asking the right questions, could be used to uncover bias in the data sets used to train the algorithms

Disinformation

The rise of disinformation, a consequence of malicious actors aiming to gain profit and power, threatens the notion of objective truth, a social foundation. With the mass availability of mobile devices and internet, digital media platforms have become influence machines of unprecedented power, qualifying anyone to broadcast to billions at almost no cost. The scale of reach, the micro-targeting techniques achieved through large data sets, and the adaptive reinforcement learning all foundational to AI technology have escalated disinformation into a current day crisis.

The consequences of disinformation are as diverse as the topics it deforms, threatening our ability to fight climate change, perpetuating human rights abuses, and undermining free and fair elections (Cadwalladr, 2017; Mozur, 2018; Hiar, 2021). Disinformation around Covid-19, for instance, has inspired widespread reservations regarding vaccines, the need for social distancing measures, and the virus' validity itself. This is further complicated by the advertising models of DPFs, the default hosts of disinformation. According to the Global Disinformation Index, platforms were inadvertently set to fund $25 million dollars of Covid-19 conspiracy theories by the end of 2020.[17] As AI becomes more powerful and widespread, the rise of malicious use is anticipated, threatening not just our social security, as disinformation does, but our political security, physical security, and digital security (Brundage et al., 2018). The issue of disinformation is therefore a challenge of governing against malicious actors; promoting economic models that disincentivize disinformation; and protecting the victims who will

[17] This estimate from this study was based on open web content only and did not include inadvertent funding on social media and video platforms (Global Disinformation Index, 2020).

experience real-life ramifications from virtual harms.

Ultimately, governing AI, and addressing challenges like bias and disinformation, will require governing the technology toolbox; and in particular, platforms.

What Should Be Governed

Debates regarding which aspects of the technology toolbox should be governed have largely focused on algorithms and data, and the ethics of use[18]. A significant tool in the box is too often overlooked in the debate about AI. Thus, rather than reviewing or extending the debates about data and algorithms, we turn our attention to platforms. Digital platforms are both a tool (i.e., platform technology) and an actor (i.e., digital platform firms), and are found throughout many, though certainly not all, of the shared concerns around AI, from data rights to algorithmic transparency.

The successful digital platform firms (DPFs) generate the pools of data from which AI draws its power; their core applications from search, through labour market matching, linking friends, and online retail—all rest on the algorithmic tools outlined in the first section of this paper. Those firms, individually in many cases, and certainly collectively, invest more in the core AI tools than nations, and are global leaders in research. Think only, for example, of Google's DeepMind. Consequently, governing AI and regulating DPFs are intimately entwined.

Platforms: The Gunslinger in the AI story

As AI is a foundational to DPF strategy, governance of AI is inextricably linked to governance of platforms. Just as novel database operations such as atomic transactions and semaphores were pivotal to enabling the rise of the time-shared client-server systems of the latter half of the 20th century, AI offers astonishingly effective techniques that enable DPFs to profit from the vast and multifarious data they collect by inferring preferences, needs, malleability, and creditworthiness of their clients and their clients' clients. In other words, AI governance will require governing DPFs. DPFs expand and amass power through big data collection and utilization, and of course, existing technology infrastructure — computing power, cloud computing, and widely available internet connectivity and smartphones. As DPFs collect data, they train and improve machine learning algorithms, thus, accumulating additional data. This cyclical relationship enables unprecedented value creation and capture and fuels profits.

Through this value creation, DPFs have emerged as a dominant business model in the 21st century, positioning them as powerful actors. When a DPF successfully enters an industry, such as Amazon in retail, Google in advertising, and Airbnb in hospitality, it becomes critical infrastructure that intermediates social and economic activity.[19] To illustrate, consider December 2020: five mega-platforms representing 1% of the 500 firms in the S&P 500 — Apple, Amazon,

[18] Zysman and Nitzberg (2020) explored the issues of data and ethics in depth.

[19] Numerous studies point to the platform-driven disruption of retail, hospitality, and an array of industries due to digital ads. See, for example, Khan, 2016; LaVecchia & Mitchell, 2016; Zervas et al., 2017; Wachsmuth and Weisler, 2018; Crain and Nadler, 2019.

Alphabet, Facebook, and Microsoft — accounted for 22% of its total worth (Wigglesworth, 2020). The rise of Amazon, from its origins in book sales to its current position in cloud computing, groceries, and even video streaming, demonstrates the ability of DPFs to expand across traditional industrial boundaries and become mega-platform firms.[20] DPFs are spreading quickly, amassing power, and rewiring the terms of competition (Kenney et al., forthcoming). Consequently, the use of AI by DPFs is central to the impact of AI. The regulation of DPFs will be central to the regulation of AI.

DPFs have, arguably, become unelected, ad hoc, private regulators in markets, and of AI in particular. They operate in an outdated regulatory environment that has not kept pace with technological change. There are consequences for market competition in many markets and for labour and labour markets. For example, as Uber emerged it challenged both regulation in transportation and rules about labour and work, particularly what has become called a gig economy. As seen in the recent California referendum,[21] regulation became salient after Uber and other ridesharing apps had already become integral businesses, enabling them to influence post hoc regulation and legitimizing that strategy.

Digital technology platforms have become central access points to critical activities from investing to communicating. Their private decisions become default regulation, shaping practices about ownership, representation, and values across the private, public, and civil sectors. Again, the governance questions, of platforms and AI, become inseparable: Having become crucial to our daily functioning, do DPFs have the unconditional ability to strip us of the access that they grant? Does our data become the default cost of use and participation in today's social and economic activities? When is the architecture and logic of an algorithm subject to transparency? Should DPFs be allowed to restructure our societies without our input on the outcome? And, of course, what are the implications of governing AI being inseparably linked to governing platforms, when we address AI governance globally?

The Complexity of Governance

We can only highlight the complexity that AI governance involves. Glimpses of that complexity and the political and social debates surrounding it can be caught from two angles.

First, AI pries open a Pandora's box of questions that sweep across the economy and society engaging diverse communities. They will not be resolved by chanting the mantras of earlier eras. Consider, that the boundaries between public and private regulation, always at issue, are blurred by DPFs. Remind ourselves that surveillance and big data question what constitutes personal privacy, privacy from corporate and government actions. Note that the easy dissemination of news, assumed for this factual news, by platforms challenges the business

[20] Kenney et al. (forthcoming) define a mega-platform firm as a firm that operates multiple platforms across industries.

[21] In November 2020, the state of California voted to approve Proposition 22, a measure allowing DPFs operating in the gig economy to continue paying workers as independent contractors. The measure, created and backed by Uber, Lyft, and DoorDash, is likely to have widespread implications for the gig economy labor laws throughout the country (Conger, 2020).

models of traditional news sources. Old questions are thus reopened for a new resolution, and new questions are posed. There is no agreement on who should regulate; on the balance between privacy and security; of what a job or protections for workers should be. Stable enduring solutions are not on the horizon, and when resolutions are generated or imposed, the new normal will likely be quite different from that of today.

Consider in one instance the complexity that each application domain represents: the problem of DPFs' viral spread of disinformation, and the range of solutions needed (Nitzberg, 2021). To begin the debate, Nitzberg argues that the range of solutions must include new types of consumer protection regulation, and enforcement mechanisms, both technology- and labour-intensive:

(1) Eliminate indemnity for platform companies in section 230 of the Communications Decency Act of 1996;
(2) Require disclosure of verifiable sources for political speech;
(3) Throttle the viral spread of messages based on verifiability of source and other measures of potential harm; and
(4) More broadly, regulate online "broadcasting" as just that; if you have 10,000 followers, you are like a small publisher.

In general, these flow from the recognition that the duty of care of a broadcaster is proportional to its reach, and that DPFs select content and route it to individuals based on their profiles learned from fine-grained data on tastes and preferences, sending just the right message at just the right times to elicit the desired response. Because this is a modified, precision-tailored 1-to-1 broadcast, it is far more effective than traditional one-to-many simultaneous broadcasting, and merits stronger, not weaker, guardrails. Disinformation illustrates the complexity of governing AI, and without even touching on competition regulation for the DPFs.

Second, the global debates mirror the domestic complexity and manifest how difficult resolutions of the issues will be. We must not exaggerate since the EU privacy initiatives did spark a complementary legislative initiative in California. But, can we imagine an agreement between China and the EU in the debate about privacy rights on the appropriate use of facial recognition tools?

Can global debates about AI move beyond ethical expressions and abstract debate? To start, nations insist they must be winners, not losers in AI development and application. AI is identified as a critical component of national success in the coming decades, as countries recognize similar opportunities and geopolitical risks posed by the suite of technologies. China's strategy[22] focuses on establishing China as an AI world leader by 2030 with investment research, promotion of applications, and indeed through international bodies defining norms and standards for AI that favour their firms. The US led strategy[23], aimed at ensuring national security and technology competitiveness can all too easily at each step define China as more than a rival.

[22] China's national AI strategy is entitled 'New Generation Artificial Intelligence Development Plan' (新一代人工智能发展规划) and was released in 2017 (Roberts et al., 2020).
[23] The White House-led strategy on AI is outlined by the National Security Commission on Artificial Intelligence (NSCAI, 2021).

Europe, not wanting to be dependent or deferrant, argues for technology sovereignty,[24] loosely pursuing a still somewhat undefined autonomy of action. In fact, over fifty countries had developed national AI strategies by the end of 2020 (HolonIQ, 2020).

If we move beyond abstract ethical expression to specific rules for particular domains and issues, there is a recognized risk in the digital marketplace. Zysman and Nitzberg (2020) argued that:

> "Indeed, more likely than agreement on global governance will be intense cyber rivalries that risk splintering the digital world into two (West vs. China), or less likely but perhaps three systems (US vs. Europe vs. China) linked but separately developing. … there will certainly be a conflict over the development of AI tools with intelligence, security, and military significance, sparking investment in AI" (p. 21).

In any case, the competing visions of what is right, what are priorities, are difficult enough to resolve domestically. Because those differences are embedded as the negotiating positions of national governments, they are all more difficult to resolve in international negotiation. To twist Robert Putnam's insight about international negotiations, there will be a two-level game, particular and general, and perhaps in multiple dimensions, local-national- and international. The two levels become intertwined of course (Putnam, 1988). For starters, consider how domestic fractions on the "best national interest" or the importance of domestic ramifications from international threats, may influence an attempt to integrate multi-national priorities. The implication we draw is that the goal must not be the ephemera of a common agreement with a single set of goals and market and social rules. Rather, one must consider how to pursue objectives of interoperability amongst nations with quite different political economies. The interpenetration of societies and economies with the fabric of digital connections is ever deeper, which makes goals of interoperability both more compelling and more difficult.

But the implications of AI suggest the need for a new stable state of global agreements which will require a fundamentally new set of global accords. If so, even interoperability may be difficult to achieve, and some fracturing of trade in services and digital flow may result. A global negotiation, such as a new round of the WTO for an AI era seems unlikely, much less succeed. That said, the recent progress toward a deal on multinational taxation and on matters such as data privacy, suggest the allies may be able to move forward in small steps that, perhaps, will accumulate into a larger deal.

Conclusion

The challenges of governing AI must be considered in the larger context of a "toolbox" including algorithms, data, processing power, and, of paramount importance, platforms (both the firms as players and technology platforms). Digital platforms generate the pools of big data on

[24] The European Commission recently released the Digital Services Act (DSA) and Digital Markets Act (DMA), which aim to strengthen the EU single market, create a safe and trustworthy online environment, and tackle issues caused by large platforms firms (Broadbent, 2020).

which AI tools operate. The regulation of digital platforms and of the data are part of the challenge of governing AI. As platforms continue to utilize AI without governance, their choices on critical matters become our default practices about ownership, representation, and values across the private, public, and civil sectors. This has escalated the need for governance of the AI and platforms. We are now in overtime.

The possibilities, applications, and risks of any such new general-purpose technologies (GPTs) only emerge over time as effects manifest on sectors of the economy and society. It is early enough in the story of AI that governance itself will determine much of the future trajectory. In the spirit of Lessig, code is its own form of law (Lessig, 2000; Lessig, 1999).

In general, governing the toolbox must balance encouraging the potential while minimizing this risk. Policy makers must well understand AI's technical aspects (e.g., its strengths, limits, and its role in the toolbox) and of the social and economic contexts in which it is situated.

AI systems already have a wide range of capabilities with vast and critical applications. But they fall short of human-level cognition and interpretation, lacking the fundamentals of context, narrative, and worldview. Today's AI refers largely to machine learning and deep learning, instruments of statistical inference building on prior data. These observations frame AI's potential, limits, and unavoidable risks of use.

New GPTs like steam power have historically seen radical economic restructuring. AI is distinct from earlier GPTs as it automates certain tasks of human cognitive capacity. This distinction, which allows AI to automate tasks in the service sector, not simply manufacturing, raises new questions about the impact on growth and labour. Still, it is policy choices that will shape the distributional consequences of this new technology toolbox, not just its capabilities in certain task areas or its pace of adoption.

The effects of AI on our communities will depend on the specific applications of AI technologies, making the consideration of community objectives and norms critical. Even if a community could agree, today, on values across an array of domains, community norms will evolve. It is important, then, to recognize the continuing evolution of social values, and the risk not only of entrenching todays' values in law, visible perhaps, but less visibly in code.

Governance must focus on sectors and applications. As the same AI application will bestow gains and inure costs distinct for each purpose and domain where it is applied, it is critical to focus primarily on sector-specific applications. Still, there will be concerns that cut across many social domains and economic sectors which are important throughout society.

AI is identified as a critical component of national success in the coming decades, as governments recognize similar opportunities and geopolitical risks posed by the suite of technologies. However, AI pries open a Pandora's box of questions that sweep across the economy and society engaging diverse communities. Moving the global debate on AI beyond ethnical expressions will therefore be unlikely. Instead, we conclude that a common agreement

around a single set of goals and market/social rules must give way to objectives of interoperability amongst nations with sometimes radically different political economies.

Works Cited

Aisch, G., Huang, J., & Kang, C. (2016). Dissecting the #PizzaGate Conspiracy Theories. *The New York Times*.

Akerlof, G., & Shiller, R. J. (2009). *Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism.* Princeton University Press.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias. *ProPublica*.

Arntz, M., Gregory, T., & Zierahn, U. (2016). *The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis* (No. 189; OECD Social, Employment and Migration Working Papers). https://doi.org/https://doi.org/10.1787/5jlz9h56dvq7-en

Bearson, D., Kenney, M., & Zysman, J. (2019). *Measuring the impacts of labor in the platform economy: new work created, old work reorganized, and value creation reconfigured.* Industrial and Corporate Change.

Black Lives Matter. (2021). *BLM Herstory*. Black Lives Matter.

Broadbent, M. (2020). *The Digital Services Act, the Digital Markets Act, and the New Competition Tool*.

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G. C., Steinhardt, J., Flynn, C., HÉigeartaigh, S. Ó., Beard, S., Belfield, H., Farquhar, S., … Amodei, D. (2018). *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*. https://arxiv.org/pdf/1802.07228.pdf

Cadwalladr, C. (2017). The great British Brexit robbery: how our democracy was hijacked. *The Guardian*.

Christian, B. (2020). *The Alignment Problem: Machine Learning and Human Values*. W. W. Norton & Company.

Cohen, P. (2016). Harold Cohen and AARON. *AI Magazine*. https://doi.org/https://doi.org/10.1609/aimag.v37i4.2695

Conger, K. (2020). Uber and Lyft Drivers in California Will Remain Contractors. *The New York Times*. https://www.nytimes.com/2020/11/04/technology/california-uber-lyft-prop-22.html

Crain, M., & Nadler, A. (2019). Political Manipulation and Internet Advertising Infrastructure. *Journal of Information Policy, 9*, 370–410.

Cutolo, D., & Kenney, M. (2020). Platform-Dependent Entrepreneurs: Power Asymmetries, Risks, and Strategies in the Platform Economy. *Academy of Management Perspectives*. https://doi.org/10.5465/amp.2019.0103

Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*.

David, P. A. (1986). *Narrow Windows, Blind Giants, and Angry Orphans: The Dynamics of Systems Rivalries and Dilemmas of Technology Policy*. Center for Economic Policy Research.

Deutsch, M. (2016). Harry Potter: Written by Artificial Intelligence. *Deep Writing*. https://medium.com/deep-writing/harry-potter-written-by-artificial-intelligence-8a9431803da6

Diaz, J. (2021). Jack Dorsey Says Trump's Twitter Ban Was "Right Decision" But Worries About Precedent. *NPR*.

Eckersley, P., & Nasser, Y. (2017). *AI Progress Measurement*. Electronic Frontier Foundation AI Progress Measurement Project.

Fatima, S., Desouza, K. C., & Gregory S. Dawson. (2020). *How different countries view artificial intelligence*. https://www.brookings.edu/research/how-different-countries-view-artificial-intelligence/

Fleischmann, K. R., & Wallace, W. A. (2006). Ethical Implications of Values Embedded in Computational Models: An Exploratory Study. *American Society for Information Science and Technology*. https://doi.org/https://doi.org/10.1002/meet.14504301254

Future of Life Institute. (2018). *AI Policy – Charlevoix Common Vision for the Future of Artificial Intelligence*. Future of Life Institute.

Gardner, H. E. (1999). *Intelligence Reframed: Multiple Intelligences for the 21st Century*. Basic Books.

Gartenberg, C. (2021). Robinhood plummets back down to a one-star rating on Google Play. *The Verge*.

Groth, O., & Nitzberg, M. (2018). *Solomon's Code: Humanity in a World of Thinking Machines*. Pegasus Books.

Hassel, A., & Sieker, F. (2020). The Effects of the Platform Economy on Work. *In 32nd Annual Meeting, SASE*.

Hiar, C. (2021). Twitter Bots Are a Major Source of Climate Disinformation. *Scientific American*.

High Level Expert Group on Artificial Intelligence. (2019). *A Definition of AI: Main Capabilities and Disciplines*. https://www.aepd.es/sites/default/files/2019-12/ai-definition.pdf

HolonIQ. (2020). *Global AI Strategy Landscape*. HolonIQ. https://www.holoniq.com/wp-content/uploads/2020/02/HolonIQ-2020-AI-Strategy-Landscape.pdf

Joshi, M., Choi, E., Weld, D. S., & Zettlemoyer, L. (2017). TriviaQA: A large scale distantly supervised challenge dataset for reading comprehension. *Association for Computational Linguistics (ACL)*, 1601–1611. https://doi.org/10.18653/v1/P17-1147

Karki, D. (2020). *Can you guess how much data is generated every day?* Takeo.

Kaushal, A., Altman, R., & Langlotz, C. (2020). Health Care AI Systems Are Biased. *Scientific American*.

Kravitz, Noah (host). *Not So Taxing: Intuit Uses AI to Make Tax Day Easier* (Episode 121). The AI Podcast.

Kelly, K., Phillips, M., & Friedman, G. (2021). Trading Curbs Reverse GameStop Rally, Angering Upstart Traders. *The New York Times*.

Kenney, M., Bearson, D., & Zysman, J. Forthcoming. *The Platform Economy Matures: Pervasive Power, Private Regulation, and Dependent Entrepreneurs.* Socio-Economic Review.

Kenney, M., & Zysman, J. (2016). *The Rise of the Platform Economy.* Issues in Science and Technology. 32(3), 61.

Kenney, M., & Zysman, J. (2019). *Unicorns, Cheshire cats, and the new dilemmas of entrepreneurial finance*. Venture Capital, 21(1), 35-50.

Khan, L. M. (2016). Amazon's Antitrust Paradox. *Yale Law Journal*, *126*(3), 710–805. https://www.yalelawjournal.org/note/amazons-antitrust-paradox

Klein, A. (2020). *Reducing bias in AI-based financial services*. https://www.brookings.edu/research/reducing-bias-in-ai-based-financial-services/

Knight, W. (2020). If AI's so smart, why can't it grasp cause and effect? *WIRED*. https://www.wired.com/ story/ai-smart-cant-grasp-cause-effect/

Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Neural Information Processing Systems, 25*. https://doi.org/10.1145/3065386

Lessig, L. (1999). *Code: And Other Laws Of Cyberspace*. Basic Books.

Liptak, A. (2021). Can Twitter Legally Bar Trump? The First Amendment Says Yes. *New York Times*. https://www.nytimes.com/2021/01/09/us/first-amendment-free-speech.html

Manyika, J., Lund, S., Chui, M., Bughin, J., Woetzel, J., Batra, P., Ko, R., & Sanghvi, S. (2017). *Jobs Lost, Jobs Gained: Workforce Transitions in a Time of Automation*.

https://www.mckinsey.com/~/media/McKinsey/Industries/Public and Social Sector/Our Insights/What the future of work will mean for jobs skills and wages/MGI-Jobs-Lost-Jobs-Gained-Executive-summary-December-6-2017.pdf

Manyika, J., Madgavkar, A., Tacke, T., Smit, S., Woetzel, J., & Abdulaal, A. (2020). *The social contract in the 21st century*. https://www.mckinsey.com/industries/public-and-social-sector/our-insights/the-social-contract-in-the-21st-century

Marcus, G. (2020). The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence. *ArXiv*, *abs/2002.0*.

Marcus, G., & Davis, E. (2019). *Rebooting AI: Building Artificial Intelligence We Can Trust*. Pantheon Books.

Marr, B. (2018). How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read. *Forbes*. https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/?sh=40c666b460ba

Marr, B. (2019). 13 Mind-Blowing Things Artificial Intelligence Can Already Do Today. *Forbes*. https://www.forbes.com/sites/bernardmarr/2019/11/11/13-mind-blowing-things-artificial-intelligence-can-already-do-today/?sh=6709939b6502

Marvin Minsky. (1968). *Semantic Information Processing*. The MIT Press.

McCarthy, J. (2007). *What is Artificial Intelligence?*

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. *AI Magazine*, 12. https://doi.org/https://doi.org/10.1609/aimag.v27i4.1904

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, *5*, 115–133.

Meltzer, J. P. (2018). *The impact of artificial intelligence on international trade*.

Mitchell, y S., & LaVecchia, O. (2016). Amazon's Stranglehold: How the Company's Tightening Grip Is Stifling Competition, Eroding Jobs, and Threatening Communities. *Institute for Local Self Reliance*. https://ilsr.org/wp-content/uploads/2020/04/ILSR_AmazonReport_final.pdf

Mozur, P. (2018). A Genocide Incited on Facebook, With Posts From Myanmar's Military. *New York Times*. https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html

National Security Commission on Artificial Intelligence. (2021). *NSCAI Final Report*. https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf

Nielsen, N. C., Murray, J., & Zysman, J. (2013). *The Services Dilemma: Productivity Sinkhole or Commoditization?* Verb Books.

Nordic Council of Ministers for Digitization. (2018). *Declaration: AI in the Nordic-Baltic region*.

OECD. (2020). *OECD Employment Outlook 2020: Worker Security and the COVID-19 Crisis*. https://doi.org/https://doi.org/10.1787/1686c758-en.

O'Neil, C. (2016). *Weapons of Math Destruction*. Crown Books.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 447–453.

Peters, J. (2021). Google salvaged Robinhood's one-star rating by deleting nearly 100,000 negative reviews. *The Verge*. https://www.theverge.com/2021/1/28/22255245/google-deleting-bad-robinhood-reviews-play-store

Rahman, K. S., & Thelen, K. (2019). The rise of the platform business model and the transformation of twenty-first-century capitalism. *Politics & Society*, *47*(2), 177–204. https://doi.org/https://doi.org/10.1177/0032329219838932

Roberts, H., Cowls, J., Morley, J., Taddeo, M., Wang, V., & Floridi, L. (2020). The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation. *AI & Society*, *36*, 59–77. https://doi.org/https://doi.org/10.1007/s00146-020-00992-2

Roose, K. (2021). In Pulling Trump's Megaphone, Twitter Shows Where Power Now Lies. *The New York Times*. https://www.nytimes.com/2021/01/09/technology/trump-twitter-ban.html

Russell, S., & Norvig, P. (2010). *Artificial Intelligence. A Modern Approach.* Pearson.

Shiller, R. J. (2019). *Narrative economics: How stories go viral and drive major economic events.* Princeton University Press.

Putnam, R. D. (1988). Diplomacy and Domestic Politics: The Logic of Two-Level Games. *International Organization*, *42*(3), 427–460.

Wachsmuth, D., & Weisler, A. (2018). Airbnb and the rent gap: Gentrification through the sharing economy. *Environment and Planning A: Economy and Space*, *50*(6), 1147–1170. https://doi.org/10.1177/0308518X18778038

Wigglesworth, R. (n.d.). How Big Tech got even bigger in the Covid-19 era. *Financial Times*. https://www.ft.com/content/d2e09235-b28e-438d-9b55-0e6bab7ac8ec

Zervas, G., Proserpio, D., & Byers, J. W. (2017). The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry. *Journal of Marketing Research*, *54*(5), 687–705.

Zysman, J. (1983). *Governments, Markets, and Growth: Financial Systems and Politics of Industrial Change*. Cornell University Press.

Zysman, J., Feldman, S., Kushida, K. E., Murray, J., & Nielsen, N. C. (2013). Services with Everything: The ICT Enabled Transformation of Services. In *The Third Globalization: Can Wealthy Nations Stay Rich in the 21st Century*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199917822.001.0001

Zysman, J., Kenney, M., & Tyson, L. D. (2019). Beyond Hype and Despair: Developing Healthy Communities in the Era of Intelligent Tools. In *INNOVATION POLICY WHITE PAPER SERIES 2019-01*. https://doi.org/10.2139/ssrn.3414691

Zysman, J., & Nitzberg, M. (2020). Governing AI: Understanding the Limits, Possibilities, and Risks of AI in an Era of Intelligent Tools and Systems. *Wilson Center*. https://www.wilsoncenter.org/sites/default/files/media/uploads/documents/WWICS Governing AI.pdf