Introduction

By Ajay Agrawal, Joshua Gans, and Avi Goldfarb

Artificial intelligence (AI) technologies have advanced rapidly over the last several years. As the technology continues to improve, it may have a substantial impact on the economy with respect to productivity, growth, inequality, market power, innovation, and employment. In 2016, the White House put out several reports emphasizing this potential impact. Despite its importance, there is little economics research on the topic. The research that exists is derived from past technologies (such as factory robots) that capture only part of the economic reach of AI. Without a better understanding of how AI might impact the economy, we cannot design policy to prepare for these changes.

To address these challenges, the National Bureau of Economic Research held its first conference on the Economics of Artificial Intelligence in September 2017 in Toronto, with support from the NBER Economics Digitization Initiative, the Sloan Foundation, the Canadian Institute for Advanced Research, and the University of Toronto's Creative Destruction Lab. The purpose of the conference was to set the research agenda for economists working on AI. The invitation emphasized these points as follows:

> "The context is this: imagine back to 1995 when the internet was about to become a thing. What would have happened to economic research into that revolution had the leading economists gathered to scope out a research agenda at that time? Today, we are facing the same opportunity with regard to AI. This time around we are convening a group of leading economists to scope out the research agenda for the next 20 years into the economics of AI."

Scholars who accepted the invitation were asked to write up and present ideas around a specific topic related to their expertise. For each paper, a discussant was assigned. Throughout the conference, in presentations, discussions, and debates, participants weighed in with their ideas for what the key questions will be, what research has already shown, and where the challenges will lie. Pioneering AI

researchers Geoffrey Hinton, Yann LeCun, and Russ Salakutinov attended, providing useful context and detail about the current and expected future capabilities of the technology. In our view, the conference was exciting because it emphasized the work that still needed to be done, rather than the presentation of standard research papers. Participants had the freedom to engage in informed speculation and healthy debate about the most important areas of inquiry.

This volume contains a summary of the proceedings of the conference. We provided authors with few constraints. This meant diversity in topics and chapter style. Many of the papers contained herein are updated versions of the original papers and presentations at the conference. Some discussants commented directly on the papers while others went further afield, emphasizing concepts that did not make it into the formal presentations but instead arose as part of debate and discussion. The volume also contains a small number of papers that were not presented at the conference, but nevertheless represent ideas that came up in the general discussion and that warranted inclusion in a volume describing the proceedings of the conference.

We categorize the papers into four broad themes. First, several papers emphasize the role of AI as a general purpose technology, building on the existing literature on general purpose technologies from the steam engine to the internet. Second, many papers that emphasize the impact of AI on growth, jobs, and inequality, focusing on research and tools from macro and labor economics. Third, six chapters discuss machine learning and economic regulation, with an emphasis on microeconomic consequences and industrial organization. The final six chapters explore how will affect research in economics.

Of course, these themes are not mutually exclusive. Discussion of AI as a GPT naturally leads to discussions of economic growth. Regulation can enhance or reduce inequality. And AI's impact on economics is a consequence of it being a general purpose technology for scientific discovery (as emphasized in chapter 6 by Cockburn, Henderson, and Stern). Furthermore, a handful of concepts cut

across the various parts, most notably the role of humans as AI improves and the interaction between technological advance and political economy.

Below, we summarize these four broad themes in detail. Before doing so, we provide a definition of the technology that makes sense for the book.

What is artificial intelligence?

The Oxford English Dictionary defines artificial intelligence as "the theory and development of computer systems able to perform tasks normally requiring human intelligence." This definition is both broad and fluid. There is an old joke among computer scientists that artificial intelligence defines what machines cannot yet do. Before a machine could beat a human expert at chess, such a win would mean artificial intelligence. After the famed match between IBM's Deep Blue and Gary Kasparov, playing chess was called computer science and other challenges became artificial intelligence.

The chapters in this volume discuss three related, but distinct, concepts of artificial intelligence. First, there is the technology that has driven the recent excitement around artificial intelligence: Machine learning. Machine learning is a branch of computational statistics. It is a tool of prediction in the statistical sense, taking information you have and using it to fill in information you do not have. Since 2012, the uses of machine learning as a prediction technology have grown substantially. One set of machine learning algorithms in particular, called "deep learning", has been shown to be useful and commercially viable for a variety of prediction tasks from search engine design to image recognition to language translation. The chapter in the book authored by us—Agrawal, Gans, and Goldfarb—emphasizes that rapid improvements in prediction technology can have a profound impact on organizations and policy (chapter 4). The chapter by Taddy (chapter 3) defines prediction with machine

learning as one component of a true artificial intelligence and provides detail on the various machine learning technologies.

While the recent interest in AI is driven by machine learning, computer scientists and philosophers have emphasized the feasibility of a true artificial general intelligence that equals or exceeds human intelligence (Bostrom 2014, Kaplan 2016). The closing sentence of this book summarizes this possibility bluntly. Daniel Kahneman writes, "I do not think that there is very much that we can do that computers will not eventually be programmed to do." The economic and societal impact of machines that surpass human intelligence would be extraordinary. Therefore—whether such an event occurs imminently, in a few decades, in a millennium, or never—it is worth exploring the economic consequences of such an event. While not a focal aspect of any chapter, several of the chapters in this volume touch on the economic consequences of such superintelligent machines.

A third type of technology that is often labeled "artificial intelligence" is better-seen as a process: automation. Much of the existing empirical work on the impact of artificial intelligence uses data on factory automation through robotics. Daron Acemoglu and Pascual Restrepo use data on factory robots to explore the impact of AI and automation on work (chapter 11). Automation is a potential consequence of artificial intelligence, rather than artificial intelligence per se. Nevertheless, discussions of the consequences of artificial intelligence and automation are tightly connected.

While most chapters in the book focus on the first definition—artificial intelligence as machine learning—a prediction technology, the economic implications of artificial general intelligence and automation receive serious attention.

AI as a GPT

A general purpose technology (GPT) is characterized by pervasive use in a wide range of sectors combined with technological dynamism (Bresnahan and Trajtenberg 1995). General purpose technologies are enabling technologies that open up new opportunities. While electric motors did reduce energy costs, the productivity impact was largely driven by increased flexibility in the design and location of factories (David 1990). Much of the interest in artificial intelligence and its impact on the economy stems from its potential as a GPT. Human intelligence is a general purpose tool. Artificial intelligence, whether defined as prediction technology, general intelligence, or automation, similarly has potential to apply across a broad range of sectors.

Brynjolfsson, Rock, and Syverson (chapter 1) argue the case for AI as a GPT. They focus on machine learning and identify a variety of sectors in which machine learning is likely to have a broad impact. They note expected continual technological progress in machine learning and a number of complementary innovations that have appeared along with machine learning. By establishing AI as a GPT, they can turn to the general lessons of the productivity literature on GPTs with respect to low rates of productivity growth, organizational challenges, and adjustment costs. Henderson (chapter 2) emphasizes the impact of a GPT on employment and the distribution of income, directly linking the discussion of AI as a GPT to questions addressed in the section on Growth, Jobs, and Inequality.

The next chapters provide micro-level detail on the nature of AI as a technology. Taddy (chapter 3) provides a broad overview of the meaning of intelligence in computer science. He then provides some technical detail on two key machine learning techniques, deep learning and reinforcement learning. Agrawal, Gans, and Goldfarb (chapter 4) argue that the current excitement around AI is driven by advances in prediction technology. They then show that modeling AI as a drop in the cost of prediction provides useful insight into the microeconomic impact of AI on organizations. They emphasize that AI is likely to substitute for human prediction, but complement other skills such as human judgment—defined as knowing the utility or valuation function. Prat's discussion (chapter 5) emphasizes that

economists typically assume that the valuation function is given, and that loosening that assumption will lead to a deeper understanding of the impact of AI on organizations.

The next chapters explore AI as a GPT that will enhance science and innovation. After reviewing the history of artificial intelligence, Cockburn, Henderson, and Stern (chapter 6) provide empirical support for the widespread application of machine learning in general, and deep learning in particular, in scientific fields outside of computer science. AI is therefore a general purpose tool in the method of invention. The paper concludes by discussing the implications for innovation policy and innovation management. Mitchell's discussion (chapter 7) emphasizes the regulatory effects of AI as a GPT for science and innovation—in terms of intellectual property, privacy, and competition policy. Several later chapters detail these and other regulatory effects. Agrawal, McHale, and Oettl (chapter 8) provide a recombinant growth model that explores how a general purpose technology for innovation could affect the rate of scientific discovery. This idea of AI as an input into innovation is a key component of several papers that appear later in the book, such as Aghion, Jones, and Jones' model of the impact of AI on economic growth (chapter 12).

The section on AI as a general purpose technology concludes with Manuel Trajtenberg's discussion of political and societal consequences (chapter 9). At the conference, Trajtenberg discussed Joel Mokyr's paper "The Past and Future of Innovation: some lessons from economic history", which will be published elsewhere. Trajtenberg's chapter does not comment directly on Mokyr, but uses Mokyr's paper as a jumping off point to discuss how technology creates winners and losers, and the policy challenges associated with the political consequences of the diffusion of a GPT. The next set of chapters also apply this idea of the emphasis on the distributional challenges of economic growth driven by rapid technological change.

Growth, jobs, and inequality

Much of the popular discussion around AI focuses on the impact on jobs. If machines can do what humans do, then will there still be work for humans in the future? The chapters in this section dig into the consequences of AI for jobs, economic growth, and inequality. Almost all chapters emphasize that technological change means an increase in wealth for society. As Jason Furman puts it in chapter 16, "We need more artificial intelligence." At the same time, several chapters emphasize that the impact of AI on society will depend on how the increased income from AI is distributed. The most recent GPTs to diffuse, computers and the internet, likely led to increased inequality due to skill-bias (e.g. Autor, Katz, and Krueger 1998; Akerman, Gaarder, and Mogstad 2015) and to an increased capital share (e.g. Autor et al 2017). This section brings together those papers that emphasized (largely macroeconomic) ideas related to growth, inequality, and jobs. If the impact of AI will be like these other technologies, what will the consequences look like for inequality, political economy, economic growth, jobs, and the meaning of work?

In chapter 10, Stevenson outlines many of the key issues. She emphasizes that economists generally agree that in the long run society will be wealthier. She highlights issues with respect to the short run and income distribution. Summarizing both the tension in the public debate and the key themes in several other chapters, she notes, "In the end, there's really two separate questions: there's an employment question, in which the fundamental question is can we find fulfilling ways to spend our time if robots take our jobs? And there's an income question, can we find a stable and fair distribution of income?"

Acemoglu and Restrepo (chapter 11) examine how AI and automation might change the nature of work. They suggest a task-based approach to understanding automation, emphasizing the relative roles of labor and capital in the economy. There is a displacement effect as machines take over human

tasks. This will lead to a lower labor share of economic output. At the same time, productivity will increase and capital will accumulate, thereby increasing the demand for labor. More importantly, growth will lead to the creation of new tasks in which labor has a comparative advantage over machines. Like Stevenson, the long run message is optimistic; however, a key theme of the paper is that adjustment costs may be high. New skills are a necessary condition of the long run optimistic forecast, and there is likely to be a short and medium term mismatch between skills and technologies. They conclude with a discussion of open questions about which skills are needed, the political economy of technological change, and the interaction between inequality and the type of innovation enabled by automation going forward.

Aghion, Jones, and Jones (chapter 12) build on the task-based model, focusing on the impact on economic growth. They emphasize Baumol's cost disease: sectors with rapid productivity growth see their share of GDP decline. Economic growth is constrained by important things that are hard to improve. This force constrains the size of the capital share with near-complete automation and even with automation of the production of ideas. Thus, even in the limit where there is an artificial general intelligence that creates a singularity or intelligence explosion with a self-improving AI, cost disease forces may constrain growth. This link between technological advance and Baumol's cost disease provides a fundamental limit to the most optimistic and the most pessimistic views. Scarcity limits both growth and the downside risk. The paper also explores how AI might reduce economic growth if it makes it easier to imitate rival's innovations. Finally, they discuss inequality within and across firms. They note that AI will increase wages of the least skilled employees of technologically advanced firms but also increasingly outsource the tasks undertaken by such employees.

Francois's discussion (chapter 13) takes this emphasis on cost disease as a starting point, asking what those tasks will be that humans are left to do. He argues that the key role for humans will be to manage the machines, even if machines are better at all tasks, humans must tell them what to optimize.

Bostrom (2014) describes this as the value-loading problem. Francois emphasizes that this is largely a political problem, and links the challenges in identifying values with Arrow's (1951) impossibility theorem. He identifies key questions around ownership of the machines, length of time that rents should accrue to those owners, and the political structure of decision-making. In raising these questions, he provides a different perspective on issues highlighted by Stevenson on the meaning of work and Trajtenberg on the political economy of technological change.

The impact on jobs has been the key focus on public discussion on AI and the economy. If human tasks get automated, what is left for humans to do? Bessen (chapter 14) explores this question, using data about other technological advances to support his arguments. He emphasizes that technological change can lead to an increase in demand and so the impact of automation on jobs is ambiguous, even within a sector.

Like Bessen, Goolsbee (chapter 15) notes that much of the popular discussion around AI relates to the labor market consequences. Recognizing that those consequences matter, his chapter mostly emphasizes the positive: Growth and productivity are good. AI has potential to increase our standard of living. Like Acemoglu and Restrepo, he notes that the short term displacement effects could be substantial. One frequently-cited solution to the displacement effects of AI is a universal basic income, in which all members of society receive a cash transfer from the government. He then discusses the economics of such a policy and the numerous challenges to making it work. Before concluding, he touches on a variety of regulatory issues that receive more detailed discussion in chapters 20 through 25. His conclusion mirrors that of Francois, emphasizing the importance of humans in determining policy direction, even if AI improves to the point where it surpasses human intelligence.

Furman (chapter 16) is similarly optimistic, emphasizing that we need more, not less AI. Referencing data on productivity growth and on the diffusion of industrial robots, he then discusses

potential negative effects on the economy as AI diffuses, particularly with respect to inequality and reduced labor force participation. Like Goolsbee, he notes several challenges to implementing a universal basic income as a solution to these negative effects. He concludes that policy has an important role to play in enabling society to fully reap the benefits of technological change while minimizing the disruptive effects.

Returning to the question of labor share highlighted by Acemoglu and Restrepo, Sachs (chapter 17) emphasizes that the income share going to capital grows with automation, and that this effect is particularly strong if human capital is counted as capital rather than labor. Furthermore, the paper argues that medium and high skilled workers will also receive a lower income share over time, as the share going to owners of capital and intellectual property grows. The chapter concludes with a list of key open questions about the dynamics of automation, the role of monopoly rents, and the consequences for income distribution and labor force participation.

Stiglitz and Korinek (chapter 18) also emphasizes income distribution, discussing the implications of AI-related innovation for inequality. They show that, in a first-best economy, contracts can be specified in advance that make innovation Pareto-improving. However, imperfect markets and costly redistribution can imply move away from the first-best. Innovation may then drive inequality directly, by giving innovators a surplus, or indirectly by changing the demand for different types of labor and capital. They discuss policies that could help reduce the increase in inequality, emphasizing different taxation tools. Stiglitz and Korinek conclude with a more speculative discussion of artificial general intelligence ("super-human artificial intelligence"), emphasizing that such a technological development will likely further increase inequality.

The final chapter in the section on growth, jobs, and inequality calls for a different emphasis. Cowen (chapter 19) emphasizes consumer surplus, international effects, and political economy. With

respect to consumer surplus, he writes, “Imagine education and manufactured goods being much cheaper because we produced them using a greater dose of smart software. The upshot is that even if a robot puts you out of a job or lowers your pay, there will be some recompense on the consumer side.” Cowen also speculates that AI might hurt developing countries much more than developed, as automation means that labor cost reasons to offshore decline. Finally, like Trajtenberg and Francois, he emphasizes the political economy of AI, highlighting questions related to income distribution.

Taken together, the chapters in this section highlight several key issues, and provide models that identify challenges related to growth, jobs, inequality, and politics. These models set up a number of theoretical and empirical questions about how AI will impact economic outcomes within and across countries.

Machine Learning and Regulation

Industry will be a key innovator and adopter of artificial intelligence. A number of regulatory issues arise. The regulatory issues related to truly intelligent machines are touched on in the chapters by Trajtenberg, Francois, Goolsbee, and Cowen. Mitchell’s discussion of Cockburn, Henderson, and Stern emphasizes intellectual property regulation. This section focuses on other regulatory challenges with respect to advances in machine learning.

Varian (chapter 20) sets up the issues by describing the key models from industrial organization that are relevant to understanding the impact of machine learning on firms. He highlights the importance of data as a scarce resource, and discusses the economics of data as an input: it is non-rival and it exhibits decreasing returns to scale in a technical sense (because prediction accuracy increases in the square root of N). He discusses the structure of ML-using industries, including vertical integration, economies of scale, and the potential for price discrimination. He concludes by highlighting policy

questions related to algorithmic collusion (which was discussed at the conference as "economist catnip," interesting and fun but unlikely to be of first-order importance), security, privacy, and transparency.

Chevalier (chapter 21) builds on Varian's emphasis on the importance of data, exploring the potential of antitrust policy aimed at companies that use machine learning. Legal scholars and policymakers have asked whether antitrust essential facilities doctrine should be applied to data ownership. She emphasizes the trade-off between static and dynamic considerations for such a policy, as well as several practical challenges.

Another regulatory issue that arises from the importance of data is privacy. Tucker (chapter 22) notes that machine learning uses data to make predictions about what individuals may desire, be influenced by, or do. She emphasizes that privacy is challenging for three reasons: cheap storage means that data may persist longer than the person who generated the data intended, non-rivalry means that data may be repurposed for uses other than originally intended, and data created by one individual may contain information about others. She discusses potential negative impacts of these three challenges, concluding with some key open questions.

Jin (chapter 23) also focuses on the importance of data as an input into machine learning. She emphasizes that reduced privacy creates security challenges, such as identify theft, ransomware, and misleading algorithms (such as Russian-sponsored posts in the 2016 U.S. election).

The economics of data also create challenges with respect to the rules governing international trade. Goldfarb and Trefler (chapter 24) argue that economies of scale in data, along with economies of scope and knowledge externalities in AI innovation, could create the opportunity for country-level rents and strategic trade policy. At the same time, they emphasize that the geographic constraints on data and knowledge would have to be high for such a policy to be optimal at the country level. They discuss

the international dimensions of domestic regulation related to privacy, access to government data, and industrial standards.

The final regulatory issue highlighted in this section is tort liability. Galasso and Luo (chapter 25) review prior literature on the relationship between liability and innovation. They emphasize the importance of getting the balance right between consumer protection and innovation incentives.

Impact on the Practice of Economics

Cockburn, Henderson, and Stern emphasize that machine learning is a general purpose technology for science and innovation. As such, it is likely to have an impact on research in a variety of disciplines, including economics. Athey (chapter 26) provides an overview of the various ways in which machine learning is likely to affect the practice of economics. She highlights that machine learning is a prediction technology, first emphasizing the usefulness of machine learning techniques for policy problems related to prediction (as in Kleinberg et al 2015). The paper then details recent advances in using machine learning techniques in causal inference, which she views as a fundamental new toolkit for empirical economists. She concludes with a list of sixteen predictions of how machine learning will impact economics, emphasizing new econometric tools, new data sets and measurement techniques, increased engagement of economists as engineers (and plumbers), and of course increased study of the economic impact of machine learning on the economy as a whole. Lederman's discussion (chapter 27) emphasizes the usefulness of machine learning to create new variables for economic analysis, and how the use of machine learning by organizations creates a new kind of endogeneity problem.

If the study of AI is going to be a key question for economists going forward, Raj and Seamans (chapter 28) emphasize that we need better data. They provide a detailed discussion of data collection opportunities by government and by academic researchers.

Related to Athey's emphasis of increased engagement of economists as engineering, Milgrom and Tadelis (chapter 29) describe how machine learning is already affecting market design decisions. Using specific examples from online marketplaces and telecommunications auctions, they emphasize the potential of AI to improve efficiency by predicting demand and supply, overcoming computational barriers, and reducing search frictions. They argue that AI will play a substantial role in the design and implementation of markets over a wide range of applications.

Camerer (chapter 30) also emphasizes the role of AI as a tool for predicting choice. He argues that firms, policymakers, and market designers can implement AI as either a "bionic patch" that improves human decision-making or "malware" that exploits human weaknesses. In addition, he explores two other ways in which AI and behavioral economics will interact. He hypothesizes that machine learning could help predict human behavior in a variety of settings including bargaining, risky choice, and games, helping to verify or reject theory. He also emphasizes that (poor) implementation of AI might provide insight into new ways to model biases in human decision-making.

The book concludes with Kahneman's brief and insightful piece (chapter 31). Kahneman begins with a discussion of Camerer's idea of using prediction to verify theory, but continues with a broader discussion of a variety of themes that arose over the course of the conference. With an optimistic tone, he emphasizes that there are no obvious limits to what artificial intelligence may be able to do.

The future of research on the economics of artificial intelligence

The chapters in this book are the beginning. They highlight key questions, recognize the usefulness of several economic models, and identify areas for further development. We can leverage what we know about GPTs to anticipate the impact of AI as it diffuses, recognizing that no two GPTs are identical. If AI is a general purpose technology, it is likely to lead to increased economic growth. A

common theme in these chapters is that slowing down scientific progress—even if it were possible—would come at a significant cost.

At the same time, many attendees emphasized that the distribution of the benefits of AI might not be even. It depends on who owns the AI, the effect on jobs, and the speed of diffusion. We can leverage what we know from growth models, skill-biased technical change, political economy, industrial organization, and econometrics to anticipate what AI will mean for particular workers, consumers, researchers, firms, and governments.

However, the technology is in its infancy. There is so much that we don't know. We hope readers of this book take it as a starting point for their own research into this new and exciting area of study.

References

Akerman, Anders, Ingvil Gaarder, and Magne Mogstad. 2015 The Skill Complementarity of Broadband Internet. *Quarterly Journal of Economics* 130(4), 1781–1824.

Arrow, Kenneth. 1951. 2nd edition 1963. Social Choice and Individual Values. John Wiley and Sons, New York.

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. 2017. The Fall of the Labor Share and the Rise of Superstar Firms. Working paper, MIT.

Autor, David H., Lawrence F. Katz, and Alan B. Krueger. 1998. Computing Inequality: Have Computers Changed the Labor Market? *Quarterly Journal of Economics*, 113(4), 1169–1213.

Bostrom, Nick. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, Oxford UK.

Bresnahan, Timothy F., and M. Trajtenberg. 1995. General purpose technologies 'Engines of growth'? *Journal of Econometrics* 65, 83-108.

David, Paul A. 1990. The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox. *American Economic Review Papers and Proceedings* 80(2), 355-361.

Kaplan, Jerry. 2016. *Artificial Intelligence: What Everyone Needs to Know*. Oxford University Press, Oxford UK.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer. 2015. "Prediction Policy Problems." *American Economic Review*, 105(5): 491-95.