1 Introduction

1.1 INTRODUCTION

Alan Turing is deservedly a hero of the modern computer age. Back in 1950 he called for the invention of thinking machines – what we would today call artificial intelligence (AI). He had a good idea how this could be done:

> We may hope that machines will eventually compete with men in all purely intellectual fields. But which are the best ones to start with? Many people think that a very abstract activity, like the playing of chess, would be best. It can also be maintained that it is best to provide the machine with the best sense organs that money can buy and teach it to understand and speak English. I think both approaches should be tried. (Turing, 1950, p. 460)

Turing argued that AI should start by playing chess and speaking English. As we are preparing this book for publication at the end of 2023, both approaches have been tried and successfully implemented. In chess, there are IBM's Deep Blue algorithm and Google's AlphaZero. The latter has taught itself to play chess in four hours and then defeated other leading chess computer programs. And as far as talking is concerned, in 2018 Google introduced Google Duplex, an AI telephone assistant that can conduct a complex conversation. Nowadays we routinely talk to our devices, whether it be Siri, Alexa, ChatGPT, or our car.

Recognizing patterns from mountains of data, whether these be chess moves or sentences, and moreover doing it better and faster than the human brain can and learning in an unsupervised manner by doing, is at the core of AI – also called *Machine Learning* (ML). It underpins a wide range of applications with which we are all to

https://doi.org/10.1017/9781009483094.002 Published online by Cambridge University Press

some extent engaging daily, whether it be filtering spam emails, performing searches on Google, engaging ChatGPT to write some code, using Bing to design an image for our next Instagram post, watching a series on Netflix, or just using our mobile devices.

While this is all very remarkable, in many ways AI seems like a typical new technology and, by that comparison, not all *that* remarkable. It has been a long way in the making. It seems to be following a typical hype cycle and *S-curve* in its reception and impact. It is, like electricity before it, slow to diffuse. Like the technologies that heralded the First Industrial Revolution, it has been associated with rising inequality – or at least the potential to widen income and wealth gaps – so nothing new there. And as with many previous technologies – for example, the steam engine, electricity, the motor car, DNA, and nuclear energy – there are hypesters and hopesters (who hype its potential benefits) and Luddites and doomsters (who lament its consequences). For example, AI has been hyped as an exponential technology by Chiacchio et al. (2018, p. 3) who claimed (with reference to a McKinsey study) that AI disruption would be 10 times faster and 300 times the scale than that of the First Industrial Revolution – thus having "3,000 times the impact." And in 2023 Yudkowsky (2023) exclaimed, "If somebody builds a too-powerful AI, under present conditions, I expect that every single member of the human species and all biological life on Earth dies shortly thereafter."

In the case of AI, the difference seems to be, at least from the present vantage point, that it is not a finished technology but a technology that is incrementally changing and still evolving. While AI currently (in its ML form) is data and energy intensive, and based on describing but not understanding intelligence, it is likely to keep on evolving. It is possible that it will become quite different in coming years. In this, AI is very different than electricity, a general-purpose technology to whom it is often compared. The scientific details of electricity are today the same as ever; it is only the way it is being engineered in applications that has differed. When the first car was

https://doi.org/10.1017/9781009483094.002 Published online by Cambridge University Press

driven out of the shop, its dangers were well known, and moreover, these have remained roughly similar. Nuclear technology, and its dangers, today is fundamentally the same as it was half a century ago. By contrast, AI is developing – it is a learning technology at the same time as humans are learning more about the nature of intelligence – and these learning processes mean that *what* precisely AI will evolve into – and when – is unknown. As Martin Rees, Astronomer Royal, pointed out, "there is no consensus among experts on the speed and advance in machine intelligence" (2018, p. 102).

This uncertainty is one reason why there is unprecedented hype – and hysteria – surrounding the technology (Naudé, 2021, 2019a). For example, how many previous technologies have spurned more than two dozen governments to formulate and adopt national strategies? How many have led calls for a specific technology to be governed from the United Nations? On an almost daily basis, new articles and position papers are being published on ethics for AI, regulation of AI, and human-centered AI. Compare this with the fact that the very real potential existential risk of bioengineered weapons and biowarfare is handled by the Biological Weapons Convention with an annual budget of US$1.4 million, as Ord (2020, p. 57) pointed out, less than that of an average McDonald's restaurant.

The uncertainty about the end game of AI and its true benefits and costs is responsible for the fascination with and horror toward AI in equal measure. For many start-up entrepreneurs, it is a potential money machine and a way to tap into the large reservoirs of venture capital that are swooshing around the world economy. Big promises can be sold on AI. For resource-starved and dying philosophy departments, the ethical dilemmas of AI has offered a new lease of life – providing a means of tapping into the funds governments feel they have to be seen spending on AI. The potential moral ambiguities and ethical traps in AI seem infinite, and philosophers have shown great creativity and are spinning out ever more thought experiments to confront us with the moral mazes of AI. For social justice warriors and Marxists, AI is a new instrument of oppression and capitalism:

https://doi.org/10.1017/9781009483094.002 Published online by Cambridge University Press

It may very well be the ultimate winner-takes-all and surveillance technology. And those – the AI doomsters – proclaiming that the end of the world is nigh, many of them funded by promoters of existential altruism, have found a new existential threat: a superintelligence that will, by definition, be an adversary that humanity can never beat.

In light of this, our book's ultimate contribution is to help reduce the uncertainty about the contemporary and future economic implications of AI. It is inspired by David Deutsch's Principle of Optimism: "If something is permitted by the laws of physics then the only thing that can prevent it from being technologically possible is not knowing how" (2011, p. 213). We need to know more about AI to clear up the uncertainties. This includes knowledge of the economics of AI and how it may shape the future economy, given that it is an evolving technology. Economics is a field that until now has only to a limited degree taken on the challenge of helping to understand AI (Agrawal et al., 2019b). This book is therefore ultimately a contribution to motivate economists to bring their insights and approaches to bear on the matter.

The scientific field of economics, which studies human exchange, has gained deep knowledge of how markets create information that coordinates decentralized decision-making toward the efficient use of resources and how societal institutions (rules of the game) affect the functioning of markets. In the case of previous technologies, economists have shown that technologies that are technically possible are often not adopted because of market-institutional features. Take, for example, the invention of farming (cultivation of crops). Before around 15,000 years ago, humans had been foragers for more than 150,000 years. The shift to farming and its technologies was always technically possible during that time, and even if humans had known *how* to farm, they would have failed to do so before markets were sufficiently large to provide economies of scale and before there were appropriate social technologies, such as property rights, to incentivize the adoption of farming (Bowles and Choi, 2019).

https://doi.org/10.1017/9781009483094.002 Published online by Cambridge University Press

We believe that economics can provide similar insights into many aspects of contemporary and future AI. This book provides illustrations of this. But the field of economics also needs to adjust its tools to be able to illuminate AI better. Key models in economics, for example, growth models, have until recently wholly abstracted from technology (and energy), focusing only on the nineteenth-century world of capital and labor as production factors. It was only in the 1990s that technology was endogenized, and the key feature of technology – as ideas that offer increasing returns and combinatorial possibilities – incorporated into economic growth models. These insights earned Paul Romer a Nobel Prize. It would be premature to imagine that the Information and communication technology (ICT) revolution, which has gathered speed only after 2007,[1] is adequately reflected in these models. We need more details and realism of digital technologies, and specifically AI, to be included in our models. These will offer gains in understanding how and why, and when, AI affects key economic outcomes such as economic growth, inequality, productivity growth, poverty, innovation and investment rates, wages, and consumption. This book illustrates how AI can be modeled in economics and how this can lead to deeper insights into this technology and reduce some of the uncertainty that surrounds it.

It is not that economists have been totally neglecting AI. The dominant approach to AI, ML, has already been used by economists since at least the 1980s to improve economic forecasting (Gogas and Papadimitriou, 2021). Economists have moreover been concerned not only about how ML can help with forecasting but also about the impacts of AI on labor markets, income distribution, innovation and productivity, allocative efficiency, competition and collusive behavior, among others – even though as we argue in this book the modeling approaches still need work. Examples of this work in economics include Acemoglu and Restrepo (2020), Aghion et al. (2017),

[1] The year 2007 was a pivotal year for the transition to the digital revolution. As Thomas Friedman memorably asked, "What the hell happened in 2007?" (2016, p. 19).

https://doi.org/10.1017/9781009483094.002 Published online by Cambridge University Press

Agrawal et al. (2019a), Berg et al. (2018), Bloom et al. (2018), Furman and Seamans (2019), Prettner and Strulik (2017), and Schiller (2019).

The concerns about AI mentioned in the previous paragraph are issues of immediate concern – topics that have also grabbed global and national political attention. The longer-term concerns, of more existential importance (Bostrom, 2014; Yudkowsky, 2008), have been neglected by economists. For instance, where will continued innovation in AI ultimately lead to? Will narrow AI make way for an artificial general intelligence (AGI)? And will this bring about continuing accelerating innovation resulting in a "Singularity"? Is superexponential, explosive, economic growth possible? Will a future AGI intentionally or unintentionally destroy humanity or, perhaps more likely, be misused by humanity? These are all questions where economists have been fairly silent, leaving the debate to be dominated by philosophers and computer scientists. Thus, this book also focuses attention on the long-term concerns about which substantial uncertainty exists, and do so from an economics viewpoint.

1.2 THE NEED FOR AN ECONOMICS OF AI

There is a strong case to be made for an economics of AI. But first, what do we mean by an economics of AI? We mean that economic tools – models – can and should be used more frequently to draw out the consequences of the development and use of AI. Such applications of AI are indeed growing. We also mean that economic models should be updated to reflect how the presence of AI affects their core assumptions. Just as like software developers issue new updates or versions of their operating systems or programs, for example, to deal with bugs or new security threats, so economists need to update their models. AI, being based in digital technologies and being a disruptor of the way information is used in economic decision-making, holds radical implications for economists' assumptions about costs, prices, competition, and distribution, among others.

Without an economics of AI, we are likely to obtain less benefit from AI and see more examples of "Awful AI" and growing fears

https://doi.org/10.1017/9781009483094.002 Published online by Cambridge University Press

of an AGI as an existential risk – with the possible unfortunate outcome that AI progress is regulated to a standstill. Most of the chapters in this book in fact illustrate this point: using existing and modified economic models to analyze the impacts of AI on the economy, the impact of policies on AI, and the economics of AI in the long run. They show that AI will neither take all our jobs nor lead to the extinction of humanity. These points can be elaborated as a way of motivating, as well as introducing, the rest of the book.

1.2.1 *The (Shorter-Run) Impacts of AI on the Economy*

A first way in which an economics of AI can help is in identifying where and how its benefits may be reduced or lost, and why and how "Awful" AI may emerge, including providing economically grounded perspectives on the realism of an AGI that may pose an existential risk.

Related to these dimensions of AI's impact is the challenge that most advanced economies face in dealing with the so-called *Great Stagnation*. So far, the impact of AI on economic and productivity growth and unemployment has been small. Even the much longer ongoing ICT revolution seems to have played out its productivity impacts (some would argue that we will have to wait a bit more time to see these, e.g., Brynjolfsson et al. [2017]). In fact a worrisome feature of the last few decades has been the stagnating productivity growth in the West. Productivity growth is, for instance, the lowest in the United Kingdom in 200 years – as Figure 1.1 shows. The sharp drop in labor productivity growth since the 1970s (the starting date of the digital/ICT "revolution") is very clear.

How – if at all – can AI reverse the Great Stagnation? And will doing so lead to massive technological unemployment?

In the first part of the book – Chapters 3 and 4 – we provide a model for analyzing the relationship between human capital and what we describe as AI abilities. The general implication of this model is to cast doubt on the likelihood that AI will lead to massive technological unemployment. It may, however, lead to higher levels

https://doi.org/10.1017/9781009483094.002 Published online by Cambridge University Press

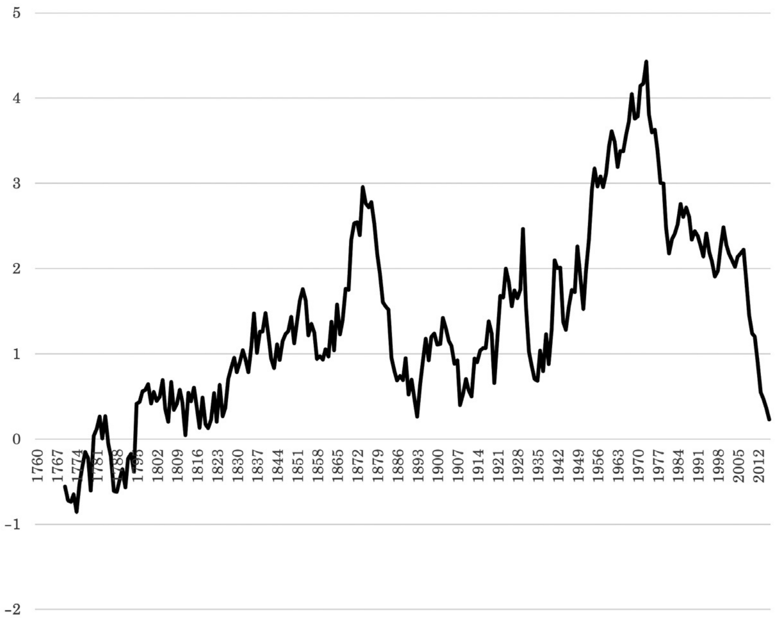

FIGURE I.1 Stagnation: Where is the impact of ICT? Labor productivity growth in the United Kingdom, 1760–2012
Source: Bank of England.

of inequality, which, through reduced aggregate demand, can lead to immiserizing growth. This will not reverse the Great Stagnation – unless, as we discuss in Chapter 10, AI develops into an innovation in the method of innovation (IMI), which can help raise research productivity. It is not clear, however, that AI will be able to fulfill these expectations. Part of the reason has to do with AI itself, and the limitations of Deep Learning. Part of the reason is outside AI: Venture capital funding to commercialize all the possible new ideas that an AI may generate will be a binding constraint.

In Chapter 10 we also consider whether AI, if it advances sufficiently, say to become an AGI, will be able to lead to accelerating economic growth – a growth explosion. Here too we tend to come to pessimistic conclusions. First, an AGI may not happen anytime soon, and if it ever does (there are reasonable grounds to assume it

https://doi.org/10.1017/9781009483094.002 Published online by Cambridge University Press

may never), it may not appear suddenly but will more likely be, as the economics of innovation suggests, the result of a long period of incremental improvements and design (as we discuss at more length in Chapter 6) – which will box in its abilities to prevent it from misalignment with human existential needs. Second, a growth explosion is likely to be very short-lived if it can overcome demand-side constraints (the topic of Chapter 5), because it will run into a brick wall of fundamental physical constraints. For instance, the energy demands from an AGI-acceleration in economic growth will quickly become prohibitive (Dutil and Dumas, 2007).

Even though an AGI may be able to increase energy efficiency and be able to decouple much growth from physical resources, it would still need significant amounts of energy to run its software and hardware – the share of the economy that can be nonphysical is ultimately bounded. If an AGI reverses the Great Stagnation and leads to a continuation of the annual average growth in energy consumption over the recent past – say the past century – of around 2,3 percent per annum, then energy use on the planet will grow from its 2019 level of 18 TW to 100 TW in 2100 and 1,000 TW in 2200. Murphy (2022b) calculates that at such a rate the economy would use up all the solar power that reaches the earth in 400 years and in 1,700 years all of the energy of the sun! The use of so much energy would generate tremendous waste heat *independent of the AGI's smart energy*. In economic modeling, the shortcomings are not taking energy as a fundamental driver of economic growth seriously and considering fundamental limits to economic growth. We discuss these shortcomings in more depth in Chapter 10.

1.2.2 The Impacts of Policy on AI

A second way in which an "economics of AI" can help is in identifying how public policy toward AI can be made better. There is indeed much enthusiasm shown by governments to implement policies to make AI more "human-centered." Much of this is unfortunately fed by the hype and hysteria that surrounds AI (for more see Chapter 2).

https://doi.org/10.1017/9781009483094.002 Published online by Cambridge University Press

Without adequate consideration of the economics of AI, governments are likely to get it wrong – making costly policy mistakes. We think that this is, unfortunately, already happening. Let us explain by discussing the current fashion for Grand National AI Strategies.

Already by 2018, at least twenty-two countries as well as the European Union had launched AI strategies, and many more announced Ethical AI frameworks. The European Union Agency for Fundamental Rights (FRA) documents more than 290 AI policy initiatives in individual EU member states between 2016 and 2020.

One country whose approach is fairly representative of these is that of Ireland. The country announced its National Artificial Intelligence Strategy, "AI – Here for Good," in July 2021. The strategy has as its ambition – similar to those of other countries' AI strategies – to see Ireland become "an international leader in using AI to benefit our economy and society, through a people-centred, ethical approach to its development, adoption and use." This is to be obtained by a comprehensive list of policy thrusts: (1) increasing trust in and understanding of AI; (2) to put appropriate governance and regulatory measures in place; (3) to promote the adoption of AI by businesses; (4) to promote the adoption of AI by the government; (5) to steer more innovation and research in AI; (6) to raise labor force skills to use and adapt to AI; and (7) to provide and secure adequate critical (ICT) infrastructure for AI systems.

Comprehensive as these national AI strategies mostly are, they tend to have shortcomings. First, they tend to uncritically share in some of the hype and hysteria surrounding AI. In the case of Ireland the AI strategy claims (p. 14) that AI could double economic growth by 2035. It fails to substantiate this by detailing critically whose growth and how.

The hype aside, a second shortcoming is that without an economic analysis of the costs, benefits, and incentives that shapes business investment and the adoption of AI, national AI strategies tend to ignore or downplay the fact that AI is central to the business models of the Chinese surveillance state and a few large digital platform firms (Google, Apple, Facebook – now Meta, Amazon,

https://doi.org/10.1017/9781009483094.002 Published online by Cambridge University Press

Alibaba – GAFAA) who enjoy winner-takes-most benefits due to network economies that characterize AI business models. These firms do not need government funding or support – in fact their research and development (R&D) budgets exceed that of many rich countries.

A third shortcoming is that, AI national strategies tend to omit consideration of the fact that it is not so much the technology per se that determines the impact but the way it transforms business models and changes the competitive landscape. AI requires large amounts of data, which in turn generate demand economies of scale, first-mover advantages, and winner-take-most effects in markets. The few companies in the world that get it right (GAFAA) become monopolists and gatekeepers, not only disrupting existing businesses but also depressing the start-up of new firms, creating virtual "killing zones" around them that stifle innovation – and which no doubt contribute to the Great Stagnation.

How to deal with this radically different (anti-) competition landscape – labelled "platform capitalism" (Srnicek, 2016), featuring platform envelopment and creative use of AI – has caused regulators and competition authorities substantial headaches (Naudé, 2023b). Not only do digital platform firms out-compete traditional "pipeline" businesses but increasingly entrepreneurs are forced to compete against each other on digital platforms – for example, on Amazon Web Services (AWS) – often with terrible results and a rise in digital subsistence entrepreneurship and destructive digital entrepreneurship (Naudé, 2023a; Van Alstyne et al., 2016). As a result, the European Union, for instance, adopted its Digital Markets Act (DMA), Digital Services Act (DSA), and AI Act in recent years to better regulate digital platforms and manage the risks posed by AI.

A fourth shortcoming of many national AI strategies is that they tend to depart from the critical assumption that there is a trust problem with AI and that this is due to people not understanding AI well enough. For example, the Irish national AI strategy aims to teaching people data science and having an AI ambassador, believing that this will raise trust (belief) in AI. In fact, from applying an economic perspective, one may expect exactly the opposite to be the

https://doi.org/10.1017/9781009483094.002 Published online by Cambridge University Press

case: The better people understand AI, the more they will see through the hype and hysteria, and the better they will realize that AI is not the panacea it is made out to be – and they will have less trust in AI.[2]

In the United States, where research and understanding of AI are quite advanced, adoption rates of AI are very low. Zolas et al. (2020) report that a 2018 US Census Bureau Survey of over 800,000 firms in the United States found that only 2,9 percent were using ML in 2018. McElheran et al. (2023) report that the adoption rate of five AI-related technologies among a sample of 850,000 US firms was, corrected for firm size, just 18 percent. A 2020 survey by the European Commission (2020) found that among EU firms who indicate using AI, "at the level of each technology, adoption in the EU is still relatively low. It ranges from merely 3% of enterprises currently having adopted sentiment analysis to 13% for anomaly detection and process/equipment optimisation." Firms do not adopt AI because it makes no business sense, not because they do not trust it. It is still just too expensive, with paltry returns for most firms, it comes with an exorbitant environmental price tag, and markets are dominated by a few incumbent firms.

Chapters 4–9 in this book provide an economic take on these issues. The analyses in the chapters show that indeed with a few firms dominating the business landscape the regulatory challenges facing governments – to for instance, incentivize human-centered AI or to limit AI arms races – may be more tractable. The analyses also highlight that to track, trace, and regulate AI advances, government regulators need to be sufficiently resourced, including having access to appropriately skilled staff, to be able to do this. In sum, the impacts of policy on AI will depend not so much on the impacts of policy on technology hardware and software but on the business models that these give rise to. Economics is well prepared to make a contribution in this regard, for instance, through insights from Game Theory, Mechanism Design, and Network Economics.

[2] This may, however, be a desirable outcome.

https://doi.org/10.1017/9781009483094.002 Published online by Cambridge University Press

1.2.3 *The Impacts of AI in the Longer Run*

> That's right: the end of the world is nigh, and it's no longer the preserve of megabudget disaster movies or bleak survivalist thrillers. These days the looming obliteration of our species can just as readily form the backdrop to some governmental mockery or a boozy country-house drama. (Hess, 2022)

The *Zeitgeist* in the third decade of the third millennium is one of *Angst*, as this quote reflects. While there have always been doomsayers predicting the imminent end of humanity, a rational, scientific approach toward understanding and acting on existential risks facing humanity is still lacking. The focus has so far largely been on measuring, mitigating, and responding to vulnerability to various idiosyncratic and covariate risks – such as risks to falling in poverty, risks to health, and the risks from natural hazards or human action – where this risk posed threats of significant damage but not to such an extent that it would "permanently or drastically curtail the potential of humanity" (Bostrom, 2002, p. 2).

However, the climate change challenge, the COVID-19 pandemic, and the renewed specter of nuclear war have made warnings that we need to face up to real existential threats more urgent. Books dealing with existential risks, including longer-term risks, have become bestsellers – see, for instance, Rees (2018), Ord (2020), and MacAskill (2022).

There is a widespread view that AI poses an existential risk. Consider, for instance, that a recent headline exclaimed that "A third of scientists working on AI say it could cause global disaster" (Hsu, 2022). According to Noy and Uher (2022, p. 498), "Artificial Intelligence (AI) systems most likely pose the highest global catastrophic and existential risk to humanity from the four risks we described here, including solar-fares and space weather, engineered and natural pandemics, and super-volcanic eruptions." AI is even seen as "millions of times more powerful than nuclear weapons" and that it "could create multiple individual global risks, most of which we can not currently imagine" (Turchin and Denkenberger, 2020, p. 148).

https://doi.org/10.1017/9781009483094.002 Published online by Cambridge University Press

In March 2023, several scientists and other notables signed an open letter published on the Future of Life Institute's web page,[3] calling "on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4. This pause should be public and verifiable, and include all key actors. If such a pause cannot be enacted quickly, governments should step in and institute a moratorium." By the time this book was going to press, this open letter has been signed by more than 33,000 people.

In Chapter 10, we argue that these fears are, just like the hopes of a Singularity and growth explosion, exaggerated. First, as we discuss in greater depth in chapter 6, an AGI is still a far way off, and if (which is a big if) it is ever invented, the process leading to its invention may very likely reduce the risks of misalignment with human objectives. Perhaps Floridi (2022, p. 9) has a point in warning that the preoccupation of certain philosophers with the Singularity and existential risks from AI "is a rich-world preoccupation likely to worry people in wealthy societies who seem to forget the real evils oppressing humanity and our planet."

However, we argue that economists ought to weigh in more on the matter of AI's potential long-term risks and discuss the reasons why they so far, have not done so. One reason is that economic risk assessment methods using expected utility theory (EUT) are not well-suited to deal with existential risks. Weitzman (2009) has proposed a dismal theorem that states that, because the probabilities of catastrophic events are characterized by long tails, EUT would assign infinite losses to it. As a result, EUT may not be able to provide an ethically acceptable approach to deal with catastrophic and existential risks. A future economics of AI may set this right.

1.3 STRUCTURE OF THIS BOOK

The rest of the book is structured as follows.

In Chapter 2, *Artificial Intelligence and Economics: A Gentle Introduction*, we describe the development of AI since World War

[3] See https://futureoflife.org/open-letter/pause-giant-ai-experiments/

https://doi.org/10.1017/9781009483094.002 Published online by Cambridge University Press

II, noting various AI "winters" and tracing the current boom in AI back to around 2006/2007. We provide various metrics describing the nature of this AI boom. We then provide a summary and discussion of the salient research relevant to the economics of AI and outline some recent theoretical advances.

Chapter 3, *Artificial Intelligence and the Economics of Decision-Making*, deals with how microeconomics can provide insights into the key challenge that AI scientists face. This challenge is to create intelligent, autonomous agents that can make rational decisions. In this challenge, they confront two questions: what decision theory to follow and how to implement it in AI systems. This chapter provides answers to these questions and makes three contributions. The first is to discuss how economic decision theory – EUT – can help AI systems with utility functions to deal with the problem of instrumental goals, the possibility of utility function instability, and coordination challenges in multi-actor and human–agent collective settings. The second contribution is to show that using EUT restricts AI systems to narrow applications, which are "small worlds" where concerns about AI alignment may lose urgency and be better labeled as safety issues. The chapter's third contribution points to several areas where economists may learn from AI scientists as they implement EUT. These include consideration of procedural rationality, overcoming computational difficulties, and understanding decision-making in disequilibrium situations.

In Chapter 4, *Artificial Intelligence in the Production Function*, the book moves from the microeconomic perspective of Chapter 3 to the macroeconomic perspective of labor markets and economic growth – although the analysis remains grounded in microeconomic functions. In this chapter, we provide an economic growth model wherein AI as a possible substitute for human labor is modeled, taking into account the nature of AI as an automation technology. This goes to the heart of the current focus of economists on AI, namely its implications for labor markets, and specifically unemployment and skills requirements. The crucial points that we make here are that economists need to go further than indirectly modeling

https://doi.org/10.1017/9781009483094.002 Published online by Cambridge University Press

AI through assumptions on substitution elasticities, and need to take the specific nature (narrow focus) of AI into explicit account.

In Chapter 5, *Artificial Intelligence, Growth, and Inequality*, we take the production function enriched with AI abilities from Chapter 4, and apply it to study the implications for progress in AI on growth and inequality. The crucial finding we discuss in this chapter is that understanding the nature of AI as narrow ML and its effect on key macroeconomic outcomes depends on having appropriate assumptions in growth models. In particular, we discuss the appropriateness of assuming, as most standard endogenous growth models today do, that economies are supply driven. If they are not supply driven, then demand constraints, which can arise from the diffusion of AI, may restrict growth. Through this we show why expectations that AI will may lead to "explosive" economic growth is unlikely to materialize: the increase in inequality and decline in consumption that will occur will act as a negative feedback effect, which will truncate the growth rate. AI progress may even contribute to negative growth. In this way, we show that by considering the nature of AI as specific (and not general) AI, and making appropriate assumptions that reflect the digital AI economy better, economic outcomes may be characterized by slow growth, rising inequality and rather full employment – conditions that rather well describe economies in the West. This chapter contributes to not only the recent theoretical literature on AI and economic growth modeling, such as the AR model, but also work by Aghion et al. (2017), Cords and Prettner (2019), Hémous and Olsen (2018), and Prettner and Strulik (2017). Unlike these models, the model presented in Chapter 5 incorporates demand constraints and a modified task approach to labor markets.

In Chapter 6, *Investing in Artificial Intelligence: Breakthroughs and Backlashes*, we move from the impacts of AI on the economy to the impacts of firm– and government-level decisions on AI. In particular, we ask what economic modeling can tell us about the likelihood that firms will invent an AGI: how much and for how long must they sustain investment in R&D to obtain such an

https://doi.org/10.1017/9781009483094.002 Published online by Cambridge University Press

invention? We develop a novel Real Options Model, one that uses a Stochastic Compound Poisson Process, to explicitly consider that a radical innovation such as an AGI is subject to much more uncertainty than typical business investments – which also helps throw light on the breakthroughs and backlashes that have characterized periodic AI winters, as is discussed in Chapter 2. The crucial insight of our model is that it will be largely government-funded agencies or state-owned enterprises efforts (e.g., by the US or Chinese governments) and/or a few large corporations (such as Google or Alibaba) that will invent an AGI, if ever. In Chapter 10, we will come back to the question of what may be the consequences if they indeed succeed.

In Chapter 7, *Artificial Intelligence Arms Races as Innovation Contests*, we go deeper into modeling one of the implications or features noted in Chapter 6, namely that a strong motivation for large firms to invest substantial amounts into R&D for an AGI is due to the winner-takes-all effects it may bestow on them. This feature, while important to incentivize AI investment, has the downside that it implies that AI arms races may take place. And the danger of an AI arms race is that it may result in an inferior AGI from a human safety perspective. In this chapter, we model such an AI arms race as an innovation contest and show how a government can steer such an arms race so as to obtain a better outcome in terms of the quality of the AGI. A crucial insight from our modeling is that the intention (or goals) of teams competing in an AGI race, as well as the possibility of an intermediate outcome ("second prize"), may be important. Making the latter available through government innovation procurement leads us to Chapter 8.

In Chapter 8, *Directing Artificial Intelligence Innovation and Diffusion*, we ask, given that Chapter 7 suggested a role for public procurement of innovation to potentially play a role in steering innovation in AI, how values and ethics in AI development can be incentivized by governments. We start out from the difficulty acknowledged in the rapidly growing field of AI ethics that the many proposals for ethical AI – or human centered AI (HCAI) – lack strong

https://doi.org/10.1017/9781009483094.002 Published online by Cambridge University Press

incentives for developers and users to adhere to them. The crucial insight from this chapter is from the use of a simple theoretical model that shows how public procurement of innovation can incentivize the development of HCAI.

Chapter 9, *Artificial Intelligence, Big Data, and Public Policy*, focuses on how public policy can steer AI, by taking how it can impact on the use of big data, one of the key inputs required for AI. Essentially, public policy can steer AI through putting conditions and limitations on data. But data itself can help improve public policy – also in the area of economic policymaking. Hence, this chapter touches on the future potential of economic policy improvements through AI. More specifically, we discuss under what conditions the availability of large data sets can support and enhance public policy effectiveness – including the use of AI – along with two main directions. We first analyze how big data can help existing policy measures to improve their effectiveness and, second, we discuss how the availability of big data can suggest new, not yet implemented, policy solutions that can improve upon existing ones. In doing so, we assume that data represent a fundamental element in policymaking. Both points are discussed within a very simple model that, despite its simplicity, provides some interesting insights. The key message of this chapter is that the desirability of big data and AI to enhance policymaking depends on the goal of public authorities and on the aspects such as the cost of data collection and storage and the complexity and importance of the policy issue.

Chapter 10, *The Future of AI and Implications for Economics*, is the final chapter. Whereas in Chapter 2, we evaluated the past and present of AI, in this chapter, we consider the future of AI. This means that unavoidably this chapter is somewhat speculative. But we build our speculation on informed discussions of the implications of current socioeconomic and technological trends and on our understanding of past digital revolutions. This allows us to provide insights on where the economy is heading and what this may imply for economics as a science.

https://doi.org/10.1017/9781009483094.002 Published online by Cambridge University Press

Future avenues for research are identified in Chapter 10. These include the need for further elaborations of economic growth models to explore the possibility of an AI-induced growth collapse, to explore the physical limits of growth, and to sharpen the tools to draw out the policy implications of facing fat-tailed catastrophic risks. Furthermore, economic perspectives may usefully be applied to the solutions and implications of the *Fermi Paradox*. These include applying economic tools to potential far-future challenges, such as decisions on whether and when – and how – to colonize the galaxy; whether or not to try and contact extraterrestrial intelligences (ETIs); whether or not to choose conflict or attempt cooperation with other ETIs; how to best protect a planetary civilization; and when an Earth-based civilization could expect to find evidence of an ETI. What Chapter 9 neatly illustrates is that delving into the economics of AI can act as a portal for economists to venture beyond the narrow confines of traditional economics – to go where no economics student has gone before.

1.4 WHO THIS BOOK IS FOR

This book is first aimed at our fellow economists – colleagues and graduate students – as a contribution to expand our field a little, and to elicit more interest for, and debate on, AI. We therefore assume that the reader of this book will be in command of a fairly high level of economic theory – especially of microeconomic optimization and economic growth theory, including familiarity with the mathematical tools – primarily calculus – that provides the language for economic theorizing. Those who are already studying economics but may not (yet) meet these requirements may benefit from first delving into some of the many great textbooks on economic growth theory. We can recommend two classics: Daron Acemoglu's *Introduction to Modern Economic Growth* (Acemoglu, 2009) and Philippe Aghion and Peter Howitt's *Endogenous Growth Theory* (Aghion et al., 1997).

We do not assume any deep level of understanding of AI models, and this is not a textbook on AI models or the application of Deep

https://doi.org/10.1017/9781009483094.002 Published online by Cambridge University Press

Learning to economics and business cases. Chapters 1–3 of the book do however provide a broad and what we consider an easy introduction into the field of AI. We do recommend however, for those not too familiar with the technical aspects of AI, the textbook of Russell and Norvig (2021), *Artificial Intelligence: A Modern Approach*, and the textbook of Deisenroth et al. (2020), *Mathematics for Machine Learning*. For those who want to jump into using ML as part of their econometric toolkit, there is Chan and Mátyás (2022)'s *Econometrics with Machine Learning*.

At this point, it is appropriate to acknowledge that this book is fundamentally theoretical in its approach. Although we do make reference to the empirical literature in economics, which has dealt mostly with the impact of AI on employment, and moreover critically reference the rather limited empirical work as stemming from the pervasive use of the task approach to labor markets (see especially Chapters 4 and 5), we see our theoretical approach as one of the strengths and unique features of the book. Our theoretical approach has two advantages: one is that it is less likely to age rapidly, unlike the case with empirical work that tends to always evolve, and which tends to be indeterminate due to difficulties comparing results from differently designed surveys, different contexts, and using different statistical methods to analyze these. A second advantage is pedagogical: the book offers a clear and consistent guide and introduction to the economics of AI. By taking a theoretical/mathematical approach, we illustrate to students and researchers new to the topic how the toolbox of economics can be applied to real-world problems.

In addition to our fellow economists, this book is also offered to fellow scientists in the fields such as (machine) ethics, philosophy, and computer science as a contribution and invitation to expand the interdisciplinary scrutinizing of AI. If economists are to do a better job at modeling AI, they would need the feedback from colleagues working in these fields. They need to understand the cutting edge of the fields of intelligence science, the science of information and of computation in particular. Hopefully, many of our colleagues in these

https://doi.org/10.1017/9781009483094.002 Published online by Cambridge University Press

fields will be able to follow our mathematical arguments, and find in them ideas that are useful for inspiring improvements in AI. As per David Deutsch's *Principle of Optimism*, it is only our limited knowledge that prevents us from designing an AI – a superintelligence – that can be of service to humanity without any of the concerns that it now raises.

https://doi.org/10.1017/9781009483094.002 Published online by Cambridge University Press