ROUTLEDGE

Why Machines Will Never Rule the World

Artificial Intelligence without Fear

JOBST LANDGREBE
AND BARRY SMITH

WHY MACHINES WILL NEVER RULE THE WORLD

The book's core argument is that an artificial intelligence that could equal or exceed human intelligence—sometimes called artificial *general* intelligence (AGI)—is for mathematical reasons impossible. It offers two specific reasons for this claim:

1. Human intelligence is a capability of a complex dynamic system—the human brain and central nervous system.
2. Systems of this sort cannot be modelled mathematically in a way that allows them to operate inside a computer.

In supporting their claim, the authors, Jobst Landgrebe and Barry Smith, marshal evidence from mathematics, physics, computer science, philosophy, linguistics, and biology, setting up their book around three central questions: What are the essential marks of human intelligence? What is it that researchers try to do when they attempt to achieve "artificial intelligence" (AI)? And why, after more than 50 years, are our most common interactions with AI, for example with our bank's computers, still so unsatisfactory?

Landgrebe and Smith show how a widespread fear about AI's potential to bring about radical changes in the nature of human beings and in the human social order is founded on an error. There is still, as they demonstrate in a final chapter, a great deal that AI can achieve which will benefit humanity. But these benefits will be achieved without the aid of systems that are more powerful than humans, which are as impossible as AI systems that are intrinsically "evil" or able to "will" a takeover of human society.

Jobst Landgrebe is a scientist and entrepreneur with a background in philosophy, mathematics, neuroscience, and bioinformatics. Landgrebe is also the founder of Cognotekt, a German AI company which has since 2013 provided working systems used by companies in areas such as insurance claims management, real estate management, and medical billing. After more than 10 years in the AI industry, he has developed an exceptional understanding of the limits and potential of AI in the future.

Barry Smith is one of the most widely cited contemporary philosophers. He has made influential contributions to the foundations of ontology and data science, especially in the biomedical domain. Most recently, his work has led to the creation of an international standard in the ontology field (ISO/IEC 21838), which is the first example of a piece of philosophy that has been subjected to the ISO standardization process.

‘It’s a highly impressive piece of work that makes a new and vital contribution to the literature on AI and AGI. The rigor and depth with which the authors make their case is compelling, and the range of disciplinary and scientific knowledge they draw upon is particularly remarkable and truly novel.’

Shannon Vallor, Edinburgh Futures Institute,
The University of Edinburgh

‘The alluring nightmare in which machines take over running the planet and humans are reduced to drudges is not just far off or improbable: the authors argue that it is mathematically impossible. While drawing on a remarkable array of disciplines for their evidence, the argument of Landgrebe and Smith is in essence simple. There can be no models and no algorithms of the complexity required to run machines which can come close to emulating human linguistic and social skills. Far from decrying AI, they laud its achievements and encourage its development; but they pour cold water on those who fail to recognise its inherent limitations. Compulsory reading for those who fear the worst, but also for those inadvertently trying to bring it about.’

Peter M. Simons FBA, Department of Philosophy
Trinity College Dublin

‘Just one year ago, Elon Musk claimed that AI will overtake humans ‘in less than five years’. Not so, say Landgrebe and Smith, who argue forcefully that it is mathematically impossible for machines to emulate the human mind. This is a timely, important, and thought-provoking contribution to the contemporary debate about AI’s consequences for the future of humanity.’

Berit Brogaard, Department of Philosophy
University of Miami

‘This book challenges much linguistically underinformed AI optimism, documenting many foundational aspects of language that are seemingly intractable to computation, including its capacity for vagueness, its interrelation with context, and its vast underpinning of implicit assumptions about physical, interpersonal, and societal phenomena.’

Len Talmy, Center for Cognitive Science
University at Buffalo

‘Interdisciplinarity is often touted, more or less successfully, as a cure for many disciplinary blind-spots. Landgrebe and Smith orchestrate a battery of arguments from philosophy, biology, computer science, linguistics, mathematics and physics in order to argue effectively and with great brio that AI has been oversold. The result is a model of how to bring together results from many different fields and argue for an important thesis. It shows that, along many dimensions, AI has not paid close enough attention to the foundations on which it rests.’

Kevin Mulligan, Department of Philosophy
University of Geneva

‘Why, after 50 years, are AI systems so bad at communicating with human beings? This book provides an entirely original answer to this question ‘AI’ but it can be summed up in just one short phrase: there is too much haphazardness in our language use. AI works by identifying patterns, for instance in dialogue, and then applying those same patterns to new dialogue. But every dialogue is different. The old patterns never work. AI must fail. If you care, read the book.’

Ernie Lepore, Department of Philosophy
Rutgers University

WHY MACHINES WILL NEVER RULE THE WORLD

Artificial Intelligence without Fear

Jobst Landgrebe and Barry Smith

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

Cover image: © Abstract Aerial Art / Getty Images

First published 2023
by Routledge
605 Third Avenue, New York, NY 10158

and by Routledge
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

Library of Congress Cataloging-in-Publication Data
A catalog record for this title has been requested

ISBN: 978-1-032-31516-4 (hbk)
ISBN: 978-1-032-30993-4 (pbk)
ISBN: 978-1-003-31010-5 (ebk)

DOI: 10.4324/9781003310105

Typeset in Bembo
by Apex CoVantage, LLC

CONTENTS

FOREWORD

Rationale for this book

This book is about artificial intelligence (AI), which we conceive as the application of mathematics to the modelling (primarily) of the functions of the human brain. We focus specifically on the question of whether modelling of this sort has limits, or whether—as proposed by the advocates of what is called the 'Singularity'—AI modelling might one day lead to an irreversible and uncontrollable explosion of ever more intelligent machines.

As concerns the current state of the art, AI researchers are, for understandable reasons, immensely proud of their amazing technical discoveries. It therefore seems obvious to all that there is an almost limitless potential for further, equally significant AI discoveries in the future.

Enormous amounts of funding are accordingly being invested in advancing the frontiers of AI in medical research, national defense, and many other areas. If our arguments hold water, then a significant fraction of this funding may be money down the drain. For this reason alone, therefore, it is probably no bad thing for the assumption of limitless potential for AI progress to be subjected to the sort of critical examination that we have here attempted.

The result, we must confess, is not always easy reading. To do our job properly, we found it necessary to move to a rather drastic degree beyond the usual disciplinary borders, drawing not merely on philosophy, mathematics, and computer science, but also on linguistics, psychology, anthropology, sociology, physics, and biology. In the "Approach" section of the Introduction we provide the rationale for this methodology and, where this is needed, for our choice of literature. In the "Glossary" (pp. 304ff.) we provide what we hope are reader-friendly definitions of the technical terms used in the main text.

We raise what we believe are powerful arguments against the possibility of engineering machines that would possess an intelligence that would equal or surpass that of humans. These arguments have immediate implications for claims, such as those of Elon Musk, according to whom AI could become 'an immortal dictator from which we would never escape'. Relax. Machines will not rule the world.

At the same time, our arguments throw light on the question of which varieties of AI *are* achievable. In this respect we are fervent optimists, and one of us is indeed contributing to the creation of new AI products being applied in industry as this book is being written. In the final chapter of the book we outline some of the positive consequences of our arguments for practical applications of AI in the future.

A new *affaire Dreyfus?*

This book is concerned not with the tremendous successes of artificial intelligence along certain narrow lanes, such as text translation or image recognition. Rather, our concern is with what is called *general* AI and with the ability of computers to emulate, and indeed to go beyond, the general intelligence manifested by humans.

We will show that it is not possible (and this means: not ever) to engineer machines with a general cognitive performance even at the level of vertebrates such as crows. When we have presented our arguments in favour of this view to friendly audiences, the most common reaction has been that we are surely just repeating the mistake of earlier technology sceptics and that our arguments, too, are doomed to be refuted by the inevitable advances of AI in the future.

Hubert Dreyfus was one of the first serious critics of AI research. His book *What Computers Can't Do*, first published in 1972, explains that symbolic (logic-based) AI, which was at that time the main paradigm in AI research, was bound to fail, because the mental processes of humans do not follow a logical pattern. As Dreyfus correctly pointed out, the logical formulation of our thoughts is merely the end-product of a tiny fraction of our mental activities—an idea which seems to be undergoing a mild revival (Fjelland 2020).

In the third edition of his book, Dreyfus (1992) was still claiming that he had been right from the beginning. And so he was; though he did not provide the sorts of arguments we give in this book, which are grounded not on Heideggerian philosophy but on the mathematical implications of the theory of complex systems.

We start out from the assumption that all complex systems are such that they obey the laws of physics. However, we then show that for mathematical reasons we cannot use these laws to analyse the behaviours of complex systems because the complexity of such systems goes beyond our mathematical modelling abilities. The human brain, certainly, is a complex system of this sort; and while there are some of today's AI proponents who believe that the currently fashionable AI paradigm of 'deep neural networks'—connectionist as opposed to symbolic AI (Minsky 1991)—can mimic the way the brain functions, we will show in what follows that, again for mathematical reasons, this is not so, not only for deep neural networks but for any other type of AI software that might be invented in the future.

We define artificial general intelligence (AGI) as an AI that has a level of intelligence that is either equivalent to or greater than that of human beings or is able to cope with problems that arise in the world that surrounds human beings with a degree of adequacy at least similar to that of human beings (a precise definition of what this means is given in sections 3.3.3–3.3.4).

Our argument can be presented here in a somewhat simplified form as follows:

A1. To build an AGI we would need technology with an intelligence that is at least comparable to that of human beings (from the definition of AGI just provided).

A2. The only way to engineer such technology is to create a software emulation of the human neurocognitive system. (Alternative strategies designed to bring about an AGI without emulating human intelligence are considered and rejected in section 3.3.3 and chapter 12.)

However,

B1. To create a software emulation of the behaviour of a system we would need to create a mathematical model of this system that enables prediction of the system's behaviour.[1]

B2. It is impossible to build mathematical models of this sort for complex systems. (This is shown in sections 8.4–8.7.)

B3. The human neurocognitive system is a complex system (see chapter 7).

B4. *Therefore*, we cannot create a software emulation of the human neurocognitive system.

From (A2.) and (B4.) it now follows that:

C. An AGI is impossible.

An analogy from physics

We conceive thesis (C.) to be analogous to the thesis that it is impossible to create a perpetual motion machine.

[1] The requirements which such predictions would need to satisfy are outlined in 3.3.4, with mathematical details in 7.1.1.4 and 8.5.

Someone might, now, argue that our current understanding of the laws of physics might one day be superseded by a new understanding, according to which a perpetual motion machine *is* possible after all.

And similarly someone might argue against the thesis of this book that our current understanding of the laws of *mathematics* might one day change. New ways of modelling complex dynamic systems may be discovered that would indeed allow the mathematical modelling of, for example, the workings of the human mind.

To see why this, too, is impossible, we show that it would have to involve discoveries even more far-reaching than the invention by Newton and Leibniz of the differential calculus. And it would require that those who have tried in the past to model complex systems mathematically, including Feynman (Feynman et al. 2010) and Heisenberg (Marshak et al. 2005, p. 76), were wrong to draw the conclusion that such an advance will never be possible. This conclusion was drawn not only by the best minds in the past. There exist also today no proposals even on the horizon of current physics or mathematics to surmount the obstacles to the modelling of complex systems identified in what follows.[2]

Acknowledgements

We thank especially Kevin Mulligan, Ingvar Johannson, and Johannes Weiß for their critical review of the manuscript. In addition, we thank Yuri Basharov, David Braun, Janna Hastings, David Hershenov, David Limbaugh, Jens Kipper, Bill Rapaport, Alan Ruttenberg, Len Talmy, Erik Thomsen, Leo Zaibert, and our anonymous reviewers for their valuable comments. We alone are responsible for any errors that remain.

The book is dedicated to our families.

[2] Two papers on *turbulence* written in 1941 by Kolmogorov (1941b, 1941a) raised some hopes that at least this complex system phenomenon could be understood mathematically. But these hopes, too, were abandoned, as we show in 8.7.1.1, with mathematical details provided in the Appendix.

1

INTRODUCTION

Since the research field of AI was first conceived in the late 1940s, the idea of an artificial general intelligence (AGI) has been put forward repeatedly. Advocates of AGI hold that it will one day be possible to build a computer that can emulate and indeed exceed all expressions of human-specific intelligence, including not only reasoning and memory, but also consciousness, including feelings and emotions, and even the will and moral thinking. This idea has been elaborated and cultivated in many different ways, but we note that AGI is far from being realised (Cohen et al. 2019).

Pennachin et al. (2007, p. 1) assert that 'AGI appears by all known science to be quite possible. Like nanotechnology, it is "merely an engineering problem", though certainly a very difficult one'. As we shall see, assertions of this sort are common in the literature on AGI. As the examination of this literature reveals, however, this is not because the thesis that there might be *fundamental* (which means: mathematical) obstacles to the achievement of AGI has been investigated and ruled out. Rather, it is because this possibility has simply been ignored.

1.1 The Singularity

Closely related to the concept of AGI is the idea of the 'Singularity', a term first applied in the AI field by Vinge (1993) and then popularised by Ray Kurzweil (2005), a pioneer of second generation AI. The term is used to refer to a point in time after which the development of AI technology becomes irreversible and uncontrollable,[1] with unknown consequences for the future of humanity.

[1] See the Glossary for a definition. The decrease and increase of the values of a function close to a singularity is hyperbolic, which is why the term 'singularity' was repurposed to describe an

DOI: 10.4324/9781003310105-1

These developments are seen by Kurzweil as an inevitable consequence of the achievement of AGI, and he too believes that we are approaching ever closer to the point where AGI will in fact be achieved (Eden et al. 2012). Proponents of the Singularity idea believe that once the Singularity is reached, AGI machines will develop their own will and begin to act autonomously, potentially detaching themselves from their human creators in ways that will threaten human civilisation (Weinstein et al. 2017). The Singularity idea features widely in debates around AGI, and it has led scientists and philosophers (as well as politicians and science fiction authors) to explore the ethical implications of this idea, for instance by postulating norms and principles that would need to be somehow built into AGIs in the future in order to counteract their potentially negative effects. Visions are projected according to which AI machines, because of their superior ethical and reasoning powers, will one day supplant existing human-based institutions such as the legal system and democratic elections. Moor (2009) talks of 'full ethical agents', which are for him the highest form of machine morality, though at the same time he believes that agents of this sort will not appear any time soon.

Visions of AGI have been associated with lavishly promoted ideas according to which we are moving towards a time when it will be possible to find cures for human diseases by applying ever more powerful computers to 'big' biological data. In the wake of the successful sequencing of the human genome and the related advent of DNA microarrays and of mass spectrometry, mass cytometry, and other sophisticated methods for performing mass assays of the fundamental components of organic matter, considerable funds have been invested in big 'ome' (transcriptome, proteome, connectome) and similar projects. Yet at the same time, even after some 20 years of research (in which both of us have participated), there is a haunting awareness of the paucity of results with significant implications for human health and disease achieved along these lines.

But we already know enough from what we have learned in the foregoing that the Singularity will not arise, given that

D1. Such a Singularity would require the engineering of an AI with the capability to engineer another machine more intelligent than itself.

D2. The exercise of this capability, at least in its early stages, would require assistance from and thus persuasive communication with human beings in bringing about the realisation of a series of highly complex goals (section 12.2.5).

D3. Only an AGI could succeed in the realisation of such goals (section 3.3).

D4. Therefore, the Singularity would require an AGI.

imagined rapid realisation of 'superintelligence' once a certain point is reached in the development of AI.

Now, however, using the proposition C (from p. xi), that an AGI is impossible and (D4.) we can infer:

E. The Singularity is impossible.

1.2 Approach

In the pages that follow we will analyse the scope and potential of AI in the future and show why the dark scenarios projected by Nick Bostrom (Bostrom 2003), Elon Musk, and others will not be realised. First, however, we set forth the sources and methods we will use. The reader interested more in content than methods may accordingly skip this section and proceed directly to page 9.

1.2.1 Realism

The overarching doctrine which binds together the different parts of this book is a non-reductivist commitment to the existence of physical, biological, social, and mental reality, combined with a realist philosophy about the world of common sense or in other words the world of 'primary theory' as expounded by the anthropologist Robin Horton (1982).[2] Thus we hold that our common-sense beliefs—to the effect that we are conscious, have minds and a will, and that we have access through perception to objects in reality—are both true and consistent with the thesis that everything in reality is subject to the laws of physics. To understand how scientific realism and common-sense realism can be reconciled, we need to take careful account of the way in which systems are determined according to the granularity of their elements. (This book is essentially a book about systems, and about how systems can be modelled scientifically.)

Central components of our realist view include the following:[3]

1. The universe consists of matter, which is made of elementary particles: quarks, leptons, and bosons.[4] Entities of various supernumerary sorts exist in

[2] Those parts of primary theory which concern human mental activities—for example thinking, believing, wanting—correspond to what elsewhere in this book we refer to as the common-sense ontology of the mental, and which is (sometimes disparagingly) referred to by analytic philosophers as 'folk psychology'.

[3] The view in question is inspired by Johansson's 'irreductive materialism' (Johansson 2012). It is similar also to the liberal naturalism expounded by De Caro (2015), which attempts 'to reconcile common sense and scientific realism in a non-Cartesian pluralist ontological perspective' and which explicitly includes as first-class entities not only material things such as you and me but also entities, such as debts, that have a history and yet are non-physical. See also Haack (2011).

[4] This is the current view, which is likely to change as physics progresses. Changes on this level will not affect any of the arguments in this book.

those parts of the universe where animals and human beings congregate. (See item 6 in this list.)

2. All interactions of matter are governed by the four fundamental forces (interactions) described by physics (electromagnetism, gravity, the strong interaction, the weak interaction), yielding all of the phenomena of nature that we perceive, including conscious human beings.
3. Fundamental entities should not be multiplied without necessity.[5] No counterpart of this maxim applies, however, to the vast realms of entities created as the products of human action and of human convention. The kilometre exists; but so also does the Arabic mile, the Burmese league, and the Mesopotamian cubit—and so do all the 'ordinary objects' discussed by Lowe (2005).
4. We thus hold that the totality of what exists can be viewed from multiple different, mutually consistent granular perspectives. From one perspective, this totality includes quarks, leptons, and the like. From another perspective it includes organisms, portions of inorganic matter, and (almost) empty space.[6]
5. At all levels we can distinguish both types and instances of these types. In addition we can distinguish at all levels continuants (such as molecules) and occurrents (such as protein folding processes).
6. Some organisms, for instance we ourselves and (we presume) you also, dear reader, are conscious. Conscious processes, which always involve an observer, are what we shall call emanations from complex systems (specifically: from organisms).[7] When viewed from the outside, they can be observed only indirectly (they can, though, be viewed directly via introspection).
7. In the world made by conscious organisms, there exist not only tapestries and cathedrals, dollar bills and drivers' licenses, but also social norms, poems, nation states, cryptocurrencies, Olympic records, mathematical equations, and data.

1.2.2 General remarks on methods

To answer the question of whether AGI is possible, we draw on results from a wide range of disciplines, in particular on technical results of mathematics and theoretical physics, on empirical results from molecular biology and other hard science domains, and (to illustrate the implications for AI of our views when

[5] This is Schaffer's Laser (Schaffer 2015).

[6] On the underlying theory of granular partitions see Bittner et al. (2001, 2003).

[7] We adapt the term 'emanation' from its usage in physics to mean any type of electromagnetic radiation, for example, thermal radiation, or other form of energy propagation (for example, sound), which is observable, but which is produced by a system process which is hidden (Parzen 2015) (cannot be observed); for a detailed definition see 2.1.

applied to the phenomenon of human conversation) on descriptive results from linguistics. In addition to the standard peer-reviewed literature, our sources in these fields include authoritative textbooks—above all the *Introduction to the Theory of Complex Systems* by Thurner et al. (2018)—and salient writings of Alan Turing, Jürgen Schmidhuber, and other leaders of AI research.

We also deal with contributions to the Singularity debate made by contemporary philosophers, above all David Chalmers (see section 9.1), and—by way of light relief—on the writings of the so-called transhumanists on the prospects for what they call 'digital immortality' (chapter 12).

1.2.3 Formal and material ontology

For reasons set forth in (Smith 2005), most leading figures in the early phases of the development of analytic philosophy adhered to an overly simplistic view of the world, which left no room for entities of many different sorts. Thus they developed assays of reality which left no room for, *inter alia*, norms, beliefs, feelings, values, claims, obligations, intentions, dispositions, capabilities, communities, societies, organisations, authority, energy, works of music, scientific theories, physical systems, events, natural kinds, and entities of many other sorts. Many analytic philosophers embraced further an overly simplistic view of the mind/brain, often taking the form of an assertion to the effect that the mind operates like (or indeed that it is itself) a computer. Computer scientists often think the opposite, namely that a computer 'acts' like the human brain and that the differences between these two types of machines will one day be overcome with the development of AGI.[8] But a computer does not *act*, and the human brain is not a machine, as we shall see in the course of this book.

In recent years, on the other hand, analytic philosophers have made considerable strides in expanding the coverage domain of their ontologies, in many cases by rediscovering ideas that had been advanced already in other traditions, as, for example, in phenomenology. They continue still, however, to resist the idea of a comprehensive realist approach to ontology. This reflects a more general view, shared by almost all analytic philosophers, to the effect that philosophy should not seek the sort of systematic and all-encompassing coverage that is characteristic of science, but rather seek point solutions to certain sorts of puzzles, often based on 'reduction' of one type of entity to another (Mulligan et al. 2006).

There is however one group of philosophers—forming what we can call the school of realist phenomenologists (Smith 1997)—who embraced this sort of comprehensive realist approach to ontology, starting out from the methodological guidelines sketched by Husserl in his *Logical Investigations* (Husserl 2000).

[8] Mathematicians who have to deal with computers take the view that computers are mere servants (*Rechenknechte*) which exist merely to perform calculations.

Like Frege, and in contrast to, for example, Heidegger or Derrida, the principal members of this school employed a clear and systematic style. This is especially true of Adolf Reinach, who anticipated in his masterwork of 1913—'The *A Priori* Foundations of the Civil Law' (Reinach 2012)—major elements of the theory of speech acts reintroduced by Austin and Searle in the 1960s.[9] Husserl's method was applied by Reinach to the ontology of law, by Roman Ingarden to the ontology of literature and music, and to the realm of human values in general (Ingarden 1986, 1973, 2013, 2019).[10]

Two other important figures of the first generation of realist phenomenologists were Max Scheler and Edith Stein, who applied this same method, respectively, to ethics and anthropology on the one hand, and to social and political ontology on the other. Ingarden is of interest also because he established a branch of realist phenomenology in Poland.[11]

The most salient members of the second generation of this school were Nicolai Hartmann, whose systematisation of Scheler's ideas on the ontology of value will concern us in chapter 6, and Arnold Gehlen, a philosopher, sociologist, and anthropologist working in the tradition of the German school of philosophical anthropology founded by Scheler.[12]

For questions of perception, person, act, will, intention, obligation, sociality, and value, accordingly, we draw on the accounts of the realist phenomenologists, especially Reinach, Scheler, and Hartmann. This is both because their ideas were groundbreaking and because their central findings remain valid today. For questions relating to human nature, psychology, and language, we use Scheler and Gehlen, though extended by the writings of J. J. Gibson and of the ecological school in psychology which he founded.

1.2.3.1 An ecological approach to mental processing

A subsidiary goal of this book is to show the relevance of environments (settings, contexts) to the understanding of human and machine behaviour, and in this we are inspired by another, less familiar branch of the already mentioned ecological

[9] (Smith 1990; Mulligan 1987) It is significant that Reinach was one of the first German philosophers to take notice, in 1914, of the work of Frege (Reinach 1969).

[10] Ingarden's massive three-volume work on formal, existential, and material ontology is only now being translated into English. This work, along with Husserl's *Logical Investigations*, provides the foundation for our treatment of the principal ontological categories (such as continuant, occurrent, role, function, and disposition) that are used in this book.

[11] One prominent member of the latter was Karol Wojtyła, himself an expert on the ethics and anthropology of Scheler (Wojtyła 1979), and it is an interesting feature of the school of phenomenological realists, perhaps especially so when we come to gauge the value of its contribution to ethics, that two of its members—namely Stein (St. Teresa Benedicta of the Cross) and Wojtyła (St. John Paul II)—were canonised.

[12] Gehlen was one of the first to explore theoretically the question of the nature and function of human language from the evolutionary perspective in his main work *Man. His Nature and Place in the World* (Gehlen 1988), first published in German in 1940 (Gehlen, 1993 [1940]).

school in psychology, which gave rise to a remarkable volume entitled *One Boy's Day: A Specimen Record of Behavior* by Barker et al. (1951). This documents over some 450 pages all the natural and social environments[13] occupied by Raymond, a typical 7-year-old schoolboy on a typical day (April 26, 1949) in a typical small Kansas town.

Barker shows how each of the acts, including the mental acts, performed by Raymond in the course of the day is tied to some specific environment, and he reminds us thereby that any system designed to emulate human mental activity inside the machine will have to include a subsystem (or better: systems) dealing with the vast and ever-changing totality of environments within which such activity may take place.

1.2.3.2 Sociology and social ontology

A further feature of the analytic tradition in philosophy was its neglect of sociality and of social interaction as a topic of philosophical concern. Matters began to change with the rediscovery of speech acts in the 1960s by Austin and Searle, a development which has in recent years given rise to a whole new sub-discipline of analytic social ontology, focusing on topics such as 'shared' or 'collective agency' (Ziv Konzelmann 2013). Many 20th-century analytic philosophers, however, have adopted an overly simplistic approach also to the phenomena of sociology and social ontology[14], and to counteract this we move once again outside the analytic mainstream, drawing first on the classical works of Max Weber and Talcott Parsons, on the writings on sociality of Gibson and his school (Heft 2017), and also on contemporary anthropologists, especially those in the tradition of Richerson et al. (2005).

1.3 Limits to the modelling of animate nature

It is well known that the utility of science depends (in increasing order) on its ability to *describe*, *explain*, and *predict* natural processes. We can *describe* the foraging behaviour of a parrot, for example, by using simple English. But to *explain* the parrot's behaviour we need something more (defined in section 7.1.1.4). And for *prediction* we need causal models, and these causal models must be mathematical.

[13] These are called 'settings' in Barker's terminology (Barker 1968), (Smith 2000). Schoggen (1989) gives an overview of the work of Barker and his school, and Heft (2001) describes the philosophical background to Barker's work and his relations to the broader community of ecological psychologists.

[14] The legal philosopher Scott Shapiro points to two major limitations of much current work on shared agency by analytic philosophers: 'first, that it applies only to ventures characterised by a rough equality of power and second, that it applies only to small-scale projects among similarly committed individuals' (Shapiro 2014). Examples mentioned by Shapiro are: singing a duet and painting a house together.

For this reason, however, it will prove that the lack of success in creating a general AI is not, as some claim, something that can and will be overcome by increasing the processing power and memory size of computers. Rather, this lack of success flows not only from a lack of understanding of many biological matters, but also from an inability to create mathematical models of how brains and other organic systems work.

In biology, valid mathematical models aiming at system explanation are hard or impossible to obtain, because there are in almost every case very large numbers of causal factors involved, making it well-nigh impossible to create models that would be predictive.[15] The lack of models is most striking in neuroscience, which is the science dealing with the physical explanation of how the brain functions in giving rise not only to consciousness, intelligence, language and social behaviour but also to neurological disorders such as dementia, schizophrenia, and autism.

The achievement of AGI would require models whose construction would presuppose solutions of some of the most intractable problems in all of science (see sections 8.5 to 8.8). It is thus disconcerting that optimism as concerns the potential of AI has been most vigorous precisely in the promotion of visions relating to enhancement, extension, and even total emulation of the human brain. We will see that such optimism rests in part on the tenacity of the view according to which the human brain is a type of computer (a view still embraced explicitly by some of our best philosophers), which on closer analysis betrays ignorance not only of the biology of the brain but also of the nature of computers. And for a further part it rests on naïve views as to the presumed powers of deep neural networks to deal with emanations from complex systems in ways which go beyond what is possible for traditional mathematical models.

1.3.1 Impossibility

Throughout this book, we will defend the thesis that it is impossible to obtain what we shall call *synoptic* and *adequate* mathematical models of complex systems, which means: models that would allow us to engineer AI systems that can fulfill the requirements such systems must satisfy if they are to emulate intelligence.

Because the proper understanding of the term *impossible* as it is used in this sentence is so important to all that follows, we start with an elucidation of how we are using it. First, we use the term in three different senses, which we refer to as *technical*, *physical*, and *mathematical* impossibility, respectively.

To say that something is *technically impossible*—for example, controlled nuclear fusion with a positive energy output—is to draw attention to the fact that it is

[15] There are important exceptions in some specific subfields, for example models of certain features of monogenetic and of infectious diseases, or of the pharmacodynamics of antibiotics. See 8.4.

impossible *given the technology we have today*. We find it useful to document the technically impossible here only where (as is all too often) proponents of transhumanist and similar concepts seek to promote their ideas on the basis of claims which are, and may for all time remain, technically impossible.[16]

To say that something is *physically impossible* is to say that it is impossible because it would contravene the laws of physics. To give an example: in highly viscous fluids (low Reynolds numbers), no type of swimming object can achieve net displacement (this is the scallop theorem [Purcell 1977]).[17]

To speak of *mathematical impossibility*, finally, is to assert that a solution to some mathematically specified problem—for example, an analytical solution of the *n*-body problem (see p. 189) or an algorithmic solution of the halting problem (see section 7.2)—cannot be found; not because of any shortcomings in the data or hardware or software or human brains, but rather for *a priori* reasons of mathematics. This is the primary sense in which we use the term *impossible* in this book.

1.4 The AI hype cycle

Despite the lack of success in brain modelling, and fired by a naïve understanding of human brain functioning, optimism as to future advances in AI feeds most conspicuously into what is now called 'transhumanism', the idea that technologies to enhance human capabilities will lead to the emergence of new 'post-human' beings. On one scenario, humans themselves will become immortal because they will be able to abandon their current biological bodies and live on, forever, in digital form. On another scenario, machines will develop their own will and subdue mankind into slavery with their superintelligence while they draw on their immortality to go forth into the galaxy and colonise space.

Speculations such as this are at the same time fascinating and disturbing. But we believe that, like some earlier pronouncements from the AI community, they must be taken with a heavy pillar of salt, for they reflect enthusiasm triggered by successes of AI research that does not factor in the fact that these successes have been achieved only along certain tightly defined paths.

In 2016 it became known that the company DeepMind had used AI in their AlphaGo automaton to partially solve the game of Go.[18] Given the complexity of the game, this must be recognised as a stunning achievement. But it is an achievement whose underlying methodology can be applied only to a narrow set of problems. What it shows is that, in certain completely rule-determined

[16] For example: 'Twenty-first-century software makes it technologically possible to separate our minds from our biological bodies.' (Rothblatt 2013, p. 317). We return to this example in chapter 11.

[17] We return to this example in section 12.3.3.1.

[18] Solving a game *fully* means 'determining the final result in a game with no mistakes made by either player'. This has been achieved for some games, but not for either chess or GO (Schaeffer et al. 2007).

confined settings with a low-dimensional phase space such as abstract games, a variant of machine learning known as reinforcement learning (see 8.6.7.3) can be used to create algorithms that outperform humans. Importantly, this is done in ways that do not rely on any knowledge on the part of the machine of the rules of the games involved. This does not, however, mean that DeepMind can 'discover' the rules governing just *any* kind of activity. DeepMind's engineers provided the software with carefully packaged examples of activity satisfying just this set of rules and allowed it to implicitly generate new playing strategies not present in the supplied examples by using purely computational adversarial settings (two algorithms playing against each other).[19] The software is not in a position to go out and identify packages of this sort on its own behalf. It is cognizant neither of the rules nor of the new strategies which we, its human observers, conceive it to be applying.

Yet the successes of DeepMind and of other AI engineering virtuosi have led once again to over-generalised claims on behalf of the machine learning approach, which gave new energy to the idea of an AI that would be *general* in the same sort of way that human intelligence is general, to the extent that it could go forth into the world unsupervised and achieve ever more startling results.

Parallel bursts of enthusiasm have arisen also in connection with the great strides made in recent years in the field of image recognition. But there are already signs that there, too, the potential of AI technology has once again been overestimated (Marcus 2018; Landgrebe et al. 2021).

Why is this so? Why, in other words, is AI once again facing a wave of dampening enthusiasm[20] representing the third major AI sobering episode after the mid-1970s and late 1980s, both of which ended in AI winters? There are, certainly, many reasons for this cyclical phenomenon. One such reason is that genuine advances in AI fall from public view as they become embedded in innumerable everyday products and services. Many contributions of working (narrow) AI are thereby hidden. But a further reason is the weak foundation of AI enthusiasm itself, which involves in each cycle an initial exaggeration of the potential of AI under the assumption that impressive success along a single front will be generalisable into diverse unrelated fields (taking us, for instance, from *Jeopardy!* to curing cancer [Strickland 2019]).[21]

[19] This is an excellent example of the use of synthetic data which is appropriate and adequate to the problem at hand.

[20] This is not yet so visible in academia and in the public prints; but it is well established among potential commercial users, for example Bloomberg is clearly indicating this in fall 2021 (https://www.bloomberg.com/opinion/articles/2021-10-04/artificial-intelligence-ain-t-that-smart-look-at-tesla-facebook-healthcare). Further documentation of a breakdown in AI enthusiasm is provided by Larson (2021, pp. 74ff.).

[21] The consequences of this assumption are thoroughly documented by Larson (2021), who explains why it is so difficult to re-engineer AI systems built for one purpose to address a different purpose.

Assumptions of this sort are made, we believe, because AI enthusiasts often do not have the interdisciplinary scientific knowledge that is needed to recognise the obstacles that will stand in the way of envisaged new AI applications. It is part of our aim here to show that crucial lessons concerning both the limits and the potential of AI can be learned through application of the right sort of interdisciplinary knowledge.

1.5 Why machines will not inherit the earth

In this book, we will argue that it is not an accident that so little progress has been made towards AGI over successive cycles of AI research. The lack of progress reflects, rather, certain narrow, structural limits to what can be achieved in this field, limits which we document in detail.

The human tendency to anthropomorphise is very powerful (Ekbia 2008). When our computer pauses while executing some especially complicated operation, we are tempted to say, 'it's thinking'. But it is not so. The processes inside the computer are physical through and through and, as we shall see in section 7.2, limited to certain narrowly defined operations defined in the 1930s by Turing and Church. The fact that we describe them in mental terms turns on the fact that the computer has been built to imitate (*inter alia*) operations that human beings would perform by using their minds. We thus impute the corresponding mental capabilities to the machine itself, as we impute happiness to a cartoon clown.

As Searle (1992) argued, computation has a physical, but no mental, reality because the significance that we impute to what the computer does (what we perceive as its mental reality) is observer dependent. If we take away all observers, then only the physical dimensions of its operations would remain. To see what is involved here, compare the difference between a dollar bill as a piece of paper and a dollar bill as money. If we take away all observers, then only the former would remain, because the latter is, again, 'observer dependent'. While we *impute* consciousness to computers, we ourselves *are* conscious.

Computers also will not be able to *gain* consciousness in the future, since as we will show, whatever remarkable feats they might be engineered to perform, the aspect of those feats we are referring to when we ascribe consciousness or mentality to the computer will remain forever a product of observer dependence.

As we discuss in more detail in chapter 9, any process that machines can execute in order to *emulate* consciousness would have to be such that the feature of consciousness that is imputed to it would be observer dependent. From the fact that a certain green piece of paper is imputed to have the observer-dependent value of one dollar, we can infer with high likelihood that this piece of paper has the value of one dollar. As we will show in chapter 9, no analogous inference is possible from the fact that a process in a machine is imputed to have the feature of consciousness (or awareness, or excitedness, or happiness, or wariness, or desire). And thus we will never be able to create an AI with the faculty of consciousness in the sense in which we understand this term when referring to humans or animals.

But if we cannot *create* consciousness in the machine, the machine might still surely be able to *emulate* consciousness? This, and the related question of the limits of computer emulation of human *intelligence*, is one of the main questions addressed in this book.

1.5.1 The nature of the human mind

As mentioned earlier, in the eyes of many philosophers working in the theory of mind, the mind works like a universal Turing machine: it takes sensory inputs and processes these to yield some behavioural output. But how does it do this? When it comes to answering this question, there are three major schools: connectionists (Elman et al. 1996), computationalists (Fodor et al. 1988), and the defenders of hybrid approaches (Garson 2018). All of them think that the mind works like a machine. Connectionists believe that the mind works like a neural network as we know it from what is called ‘artificial neural network’ research.[22] Computationalists believe that the mind operates by performing purely formal operations on symbols.[23]

Most important for us here are the hybrid approaches as pioneered by Smolensky (1990), which seek to reconcile both schools by proposing that neural networks can themselves be implemented by universal Turing machines, an idea that was indeed technically realised in 2014 by Graves et al. (2014), who used a neural network to implement a classical von Neumann architecture. Their result proves that the initial dispute between connectionists and computationalists was mathematically nonsensical, because it shows that a universal Turing machine can implement *both* connectionist *and* symbolic logic. In other words, both types of computational procedures can be expressed using the basic recursive functions defined by Alonzo Church (1936). That both symbolic and perceptron (neural network) logic are Turing-computable has been known to mathematicians and computer scientists since the 1950s, and this makes the whole debate look naïve at best.

However there is a deeper problem with all ideas according to which the functioning of the mind (or brain) can be understood and modelled as the functioning of one or other type of machine, namely that such ideas are completely detached from the standpoint of biology and physics.[24] We will show that the

[22] An artificial neural network is an implicit mathematical model generated by constraining an optimisation algorithm using training data and optimisation parameters; see further in chapter 8.

[23] The relation between these two schools from an AI research perspective is summarised by Minsky (1991), who made important contributions to connectionist AI.

[24] We shall see in detail why this is so in chapter 2 and section 9.4 and 12.2.

mentioned alternatives fail, because the mind (or brain) does not operate like a machine, and those who propose that it does do not acknowledge the results of neuroscience. For while we do not know how the mind works exactly, what we do know from neuroscience is that the workings of the mind resist mathematical modelling.[25] Therefore, we cannot emulate the mind using a machine, nor can we engineer other non-machine kinds of complex systems to obtain so-far undescribed kinds of non-human intelligence, and we will understand why in the course of this book.

1.5.2 There will be no AGI

The aim of AGI research, and of those who fund it, is to obtain something useful, and this will imply that an AGI needs to fulfill certain requirements—broadly, that it is able to cope with the reality in which humans live with a level of competence that is at least equivalent to that of human beings (see sections 3.3.3 and 3.3.4). We show that this is not possible, because there is an upper bound to what can be processed by machines. This boundary is set, not by technical limitations of computers, but rather by the limits on the possibilities for mathematical modelling.

There can be no 'artificial general intelligence', and therefore no 'Singularity' and no 'full ethical agents', because all of these would lie way beyond the boundary of what is even in principle achievable by means of a machine.

As we show at length in chapters 7 and 8, this upper bound is not a matter of computer storage or processing speed—factors which may perhaps continue to advance impressively from generation to generation. Rather, it is a matter of mathematics, at least given the presupposition that the aim is to realise AGI using computers.[26] For every computational AI is, after all, just a set of Turing-computable mathematical models taking input and computing output in a deterministic manner. Even 'self-learning' stochastic models behave deterministically once they have been trained and deployed to operate in a computer.

We shall deal in this book with all types of models that can currently be used to create computer-based AI systems, and we present each in great detail in chapter 8. Our arguments are completely general; they apply to all these types of models in the same way, and we are confident that these same arguments will apply also to any new types of models that will be developed in the future. At the same time, however, we note that these arguments potentially provide a boon to our adversaries, who can use them as a guide to the sorts of obstacles that would need to be overcome in order to engineer an AI that is both more useful than what we already have, and feasible from the point of view of engineering.

[25] The 1,696 pages of *Principles of Neural Science* by Kandel et al. (2021), which is the gold standard textbook in the field and summarises some 100 years of neuroscientific research, contain almost no mathematics. And this is not about to change.

[26] Other approaches, for example resting on the surgical enhancement of human brains, are considered in section 12.2.4.

1.5.3 Prior arguments against artificial human-level intelligence

We are not alone in believing that the idea of AGI, and of the Singularity which will follow in its wake, is at least to some degree a reflection of overconfidence among some members of the AI research community, and a number of AI proponents have expressed views which anticipate at least part of what we have to say here. For example, and most usefully, Walsh (2017). Walsh does indeed believe that AI with human-level intelligence will be achieved within the next 30–40 years; but he holds at the same time that there are a number of reasons why the Singularity will not arise:

1. intelligence is much more than thinking faster,
2. humans may not be intelligent enough to design superintelligence,
3. there is no evidence at all that an ML (machine learning) algorithm which achieves human level intelligence would thereby somehow proceed to becoming *more* intelligent (what David Chalmers [2010] calls 'AI+'),
4. there are diminishing returns from AI performance, so that performance improvements to the level of a Singularity may be stymied,
5. systems have physical limits, and there are 'empirical laws that can be observed emerging out of complex systems'. Intelligence itself as 'a complex phenomenon may also have such limits that emerge from this complexity. Any improvements in machine intelligence, whether it runs away or happens more slowly, may run into such limits' (op. cit., p. 61)[27],
6. the computational complexity required to go beyond human level intelligence may not be physically realisable.

Other important reservations concerning the possibility of the Singularity and the limits of AI in general have been brought forward by:

- Yann LeCun, who addresses the claims made by some researchers concerning an anticipated exponential growth in the powers of AI and points out that,

  > the first part of a sigmoid looks a lot like an exponential. It's another way of saying that what currently looks like exponential progress is very likely to hit some limit—physical, economical, societal—then go

[27] We note in passing that this may be one reason for the apparent contradiction between the lack of evidence for extraterrestrial civilisations and various high estimates for their probability (Fermi's paradox). Why do we see no evidence of alien superintelligences? Because the same limits to the increase in power of AI would (we believe) apply also to any technology developed by other intelligent life forms. This has implications also for the idea, favoured by Elon Musk, according to which the world in which we live is a simulation.

> through an inflection point, and then saturate. I'm an optimist, but I'm also a realist.
>
> (LeCun 2015)

- Yoshua Bengio, who makes the point that it is impossible to teach machines moral judgement: 'People need to understand that current AI—and the AI that we can foresee in the reasonable future—does not, and will not, have a moral sense or moral understanding of what is right and what is wrong' (Ford 2018, p. 31).
- Judea Pearl, who emphasises that the currently fashionable stochastics-based 'opaque learning machines' (Pearl 2020) lack an important feature of human-level intelligence in that they cannot answer questions related to causality and thus they cannot develop understanding about how things work. Pearl does not exclude the possibility of creating an AGI. He insists only that 'human-level AI cannot emerge solely from model-blind learning machines; it requires the symbiotic collaboration of data and models'.[28]
- Brian Cantwell Smith (2019), who states that

> neither deep learning nor other forms of second-wave AI, nor any proposals yet advanced for third-wave, will lead to genuine intelligence. Systems currently being imagined will achieve formidable reckoning prowess, but human-level intelligence and judgment, honed over millennia, is of a different order.
>
> (*The Promise of Artificial Intelligence*, Introduction)

- For Shannon Vallor:

> Those who are predicting an imminent 'rise of the robots' or an 'AI singularity,' in which artificially intelligent beings decide to dispense with humanity or enslave us, in my view serve as an unhelpful distraction from the far more plausible but less cinematic dangers of artificial intelligence. These mostly involve unexpected interactions between people and software systems that aren't smart enough to avoid wreaking havoc on complex human institutions, rather than robot overlords with 'superintelligence' dwarfing our own.
>
> (Vallor 2016, p. 250)

- Steven Pinker argues that the threats to freedom in the future lie not so much in the advent of any putative Singularity, but rather in the way

[28] We shall see what this means in chapter 8; essentially, that the AI we can realise is determined by us.

societies choose to use technology. He draws what we shall recognise later as the crucial distinction between intelligence and motivation. And while he is ready to accept that we might technically realise something like the former, he points out that 'there is no law of complex systems that says that intelligent agents must turn into ruthless megalomaniacs'. He also clearly sees that intelligence is not a boundless continuum with no limits to its potency (a point which we discuss in chapter 12); he recognises that stochastic models do not create knowledge and that AI is just a technology like any other, which is 'constantly tweaked for efficacy and safety' (Pinker 2020).

- Darwiche stated in (2018) that

> what just happened in AI is nothing close to a breakthrough that justifies worrying about doomsday scenarios.... The current negative discussions by the public on the AI Singularity, also called super intelligence, can only be attributed to the lack of accurate ... characterisations of recent progress. (p. 66)

Although such expressions of AGI pessimism are rarely encountered in the public prints, we suspect that the passages just cited in fact represent the views of a majority of AI experts. But they are all arguments to the effect that the Singularity *might not happen*. Here, in contrast, we will present arguments to the effect that already the creation of AI with an intelligence comparable to that of a human being is *impossible to achieve*, and thus that the Singularity, too, *will never happen*.

One notable exception is François Chollet, who argues that the idea of an 'intelligence explosion comes from a profound misunderstanding of both the nature of intelligence and the behavior of recursively self-augmenting systems'.[29] His main hypotheses are:

- that AGI theorists employ an erroneous definition of intelligence,
- that human intelligence depends on innate dispositions, on interaction with the environment (sensorimotor affordances), and on socialisation; it can be exemplified only by a human being who is part of a society,
- that complex real-world systems cannot be modelled using the Markov assumption.

Chollet points out further that the 'no free lunch theorem' (8.6.6.3) implies that if 'intelligence is a problem-solving algorithm, then it can only be understood with respect to a specific problem'. In sum, Chollett defends a view of AGI very much in the spirit of this book, but he provides only limited arguments on behalf of this view, as contrasted with the sort of detailed discussion that we present in

[29] Retrieved at https://medium.com/@francois.chollet/the-impossibility-of-intelligence-explosion-5be4a9eda6ec

chapters 7 and 8. For we will demonstrate that it is impossible to create the sorts of mathematical models even of vertebrate intelligence that would be needed in order to engineer its counterpart in a computer.

1.5.3.1 On abduction

A more recent, and for our purposes more significant, contribution to the debate on AGI is the book by Larson (2021). Larson hedges his bets as to whether human-level AI will or will not be achieved in the future, though he points out that 'no one has the slightest clue how to build an artificial general intelligence' (p. 275). But he emphasises that we do not have today, even on the horizon, anything like human-level AI. This is so, he argues, because of the current dominance of the assumption that the arrival of AGI is only a matter of time, because 'we have already embarked on the path that will lead to human-level AI, and then superintelligence'. He calls this assumption 'the myth of AI', arguing that the assumption of inevitability is so deeply entrenched that—as we ourselves have discovered in many of our encounters with AI scientists—arguing against it is taken as a form of Luddism. Larson points out in this connection that there are after all strong incentives for proponents of AI to keep its limitations in the dark, where a healthy culture for innovation 'emphasises exploring unknowns, not hyping extensions of existing methods—especially when these methods have been shown to be inadequate to take us much further'.

As we shall see in great detail in what follows, human and machine intelligence are radically different. The myth of AI insists that the differences are only temporary, in the sense that, step-by-step, more powerful AI systems will erase them. Yet, as Larson points out, the success achieved by focusing on narrow AI applications such as game-playing or protein folding 'gets us not one step closer to general intelligence. … No algorithm exists for general intelligence. And we have good reason to be skeptical that such an algorithm will emerge through further efforts on deep learning systems or any other approach popular today'. To identify one potential alternative approach, Larson points to what he sees as the three different types of inference: *deduction*, which is explored by classic symbolic AI; *induction*, which he classifies as the province of modern stochastic AI[30]; and a third type which, following the American pragmatist philosopher Peirce, he calls *abduction*. Peirce's term is nowadays used in different contexts as another word for 'hypothesis formation' or also just plain 'guessing'.[31]

[30] This is not correct, as we shall see in chapter 8.6.6.1. Machines do not engage in inductive reasoning; they rather compute local minima for loss functions, which can be seen as a very primitive emulation of induction from data because a functional is indeed obtained from observations (individual data). However, machines do not perform the induction themselves; they merely compute human-designed optimisation algorithms which emulate a narrow form of human induction.

[31] For an account of problems we might face in formalising Peirce's notion, see Frankfurt (1958).

It is abduction, Larson argues, which is at the core of human intelligence, and thus engineering a counterpart of abduction—a combination of intuition and guessing—would be needed for human-level AI. His book provides a thorough and convincing account of why this is so. Yet at the same time he complains that 'no one is working on it—at all'.

His explanation for this lack of interest is that the myth of AI is holding back AI researchers. Yet this surely underestimates the degree to which the AI field is and has always been unrestrainedly opportunistic. For if modeling abduction truly provided even the beginnings of a feasible path toward modeling human-level intelligence, would there not be contrarian AI researchers who would have started off already down this path?

The fact, if it is a fact, that there is no one who is exploring a strategy along these lines leads us to postulate that this is not for reasons having to do with the culture of AI research. Rather, it is because attempts to engineer the types of abductive inference characteristic of human reasoning have in every case failed to reach even first base. The reasons for this are explored in what follows. For where, already on the first page of his book, Larson asserts that 'the future of AI is a scientific unknown', we show that there are in fact many things that we know about the future of AI, all of which derive from the premise that any AI algorithm must be Church-Turing computable.

1.6 How to read this book

If you have not done so already, please go back and read the Foreword. This provides an account of how the AGI scepticism defended in this, book differs from earlier varieties of scepticism, in that it is based in mathematics, physics, and biology.

For those who want to go straight to the technical details of our argument against the possibility of AGI, read chapters 7 and 8 first. The earlier chapters are there to set the scene, especially as concerns the reasons why human dialogue and human ethics cannot be modelled in a neural network because of the impossibility of collecting representative samples that can be used for training.

For everyone else: read chapter 2 to understand our view of the relationship between the mental and the physical: mental events are a special type of physical event in the brain and are subject to the same laws. We argue that this view is consistent with a common-sense understanding of human mental activity (of how it feels from the inside to be a conscious human being).

Read chapter 3 to understand what the 'intelligence' is that AI researchers are seeking to emulate. We introduce a distinction between two types of intelligence: the basic kind, which we share with higher animals; and the type of intelligence that is unique to humans and is closely associated with our ability to use language. We then examine the definition of 'intelligence' used by AI researchers and show that this definition does justice to neither of these.

Read chapters 4 to 6 to get an idea of the complex systems formed when human beings interact. These chapters survey our social capabilities as humans, including our capability to use language, to follow social (including ethical) norms, and to engage in social interactions. Human languages and human societies are complex systems—in fact they are complex systems of complex systems.

Read chapters 7 and 8 to understand what complex systems are and why their behaviour cannot be modelled mathematically and therefore cannot be emulated by using computers. We survey attempts to model complex systems in medicine, psychology, and economics. We survey the entire mathematical repertoire of available approaches to the emulation of complex systems, from recursive neural networks through evolutionary process models to entropy models. And we show why they all fail.

Read chapter 9 to find out how the results obtained so far throw light on philosophers' attempts to demonstrate that an AGI, and with it the 'Singularity', can be achieved 'before long'. We focus especially on the attempts by David Chalmers to show how the Singularity might be achieved, either by emulating human intelligence in a machine or by creating a machine intelligence that would emulate the entire course of evolution.

Read chapters 10 and 11 to find out why machines cannot emulate human conversation or moral behaviour. We cover in detail why machines will neither conduct conversations nor interpret text as humans can for a variety of reasons again having to do with the properties of complex systems. We then show why this same complexity rules out the possibility of an AI ethics.

Read chapter 12 if you are interested in 'transhumanism' and in what some are pleased to call 'digital immortality'. Here we address some of the more outlandish speculations that have grown up in the hinterlands of the Singularity. We demonstrate that we can neither create a machine emulation of anything like the human mind nor transcend our human condition as mortal organisms with organic bodies in order to enjoy immortal life in digital form. We also show, along the way, that there will be no AGI, and no Singularity.

Read chapter 13 if you are interested in what can still be achieved by AI in the future, even after taking account of the limits identified in this book. For there are still many grounds for optimism as concerns the potential uses of AI. This chapter is entitled 'AI spring eternal', and it describes how narrow AI will intensify and further broaden the technosphere that mankind has been creating since the beginning of urbanisation and the advent of the first high cultures. Even though there will be no AGI, and no Singularity, AI in the narrow sense will prove itself able to bring about new and still unconceived enhancements and extensions to the texture of our industrialised societies.