Margaret A. Boden

ARTIFICIAL INTELLIGENCE

A Very Short Introduction

OXFORD

Margaret A. Boden

ARTIFICIAL INTELLIGENCE

A Very Short Introduction

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide. Oxford is a registered trade mark of Oxford University Press in the UK and in certain other countries

First published in hardback as *AI: Its Nature and Future* 2016
First published as a *Very Short Introduction* 2018

Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2018940942

ISBN 978–0–19–960291–9

Printed in Great Britain by
Ashford Colour Press Ltd, Gosport, Hampshire

Contents

Acknowledgements

I thank the following friends for their very helpful advice (any mistakes, of course, are mine): Phil Husbands, Jeremy Reffin, Anil Seth, Aaron Sloman, and Blay Whitby. And I thank Latha Menon for her understanding and patience.

List of illustrations

Chapter 1
What is artificial intelligence?

Artificial intelligence (AI) seeks to make computers do the sorts of things that minds can do.

Some of these (e.g. reasoning) are normally described as 'intelligent'. Others (e.g. vision) aren't. But all involve psychological skills—such as perception, association, prediction, planning, motor control—that enable humans and animals to attain their goals.

Intelligence isn't a single dimension, but a richly structured space of diverse information-processing capacities. Accordingly, AI uses many different techniques, addressing many different tasks.

And it's everywhere.

AI's practical applications are found in the home, the car (and the driverless car), the office, the bank, the hospital, the sky…and the Internet, including the Internet of Things (which connects the ever-multiplying physical sensors in our gadgets, clothes, and environments). Some lie outside our planet: robots sent to the Moon and Mars, or satellites orbiting in space. Hollywood animations, video and computer games, sat-nav systems, and Google's search engine are all based on AI techniques. So are the systems used by financiers to predict movements on the stock market, and by national governments to help guide policy decisions in health and transport. So are the apps on mobile phones. Add avatars in virtual reality, and the toe-in-the-water models of emotion developed for 'companion' robots. Even art galleries use AI—on their websites, and also in exhibitions of computer art. Less happily, military drones roam today's battlefields—but, thankfully, robot minesweepers do so too.

AI has two main aims. One is *technological*: using computers to get useful things done (sometimes by employing methods very *unlike* those used by minds). The other is *scientific*: using AI concepts and models to help answer questions about human beings and other living things. Most AI workers focus on only one of these, but some consider both.

Besides providing countless technological gizmos, AI has deeply influenced the life sciences.

In particular, AI has enabled psychologists and neuroscientists to develop powerful theories of the mind–brain. These include models of *how the physical brain works*, and—a different, but equally important, question—*just what it is that the brain is doing*: what computational (psychological) questions it is answering, and what sorts of information processing enable it to do that. Many unanswered questions remain, for AI itself has taught us that our minds are very much richer than psychologists had previously imagined.

Biologists, too, have used AI—in the form of 'artificial life' (A-Life), which develops computer models of differing aspects of living organisms. This helps them to explain various types of animal behaviour, the development of bodily form, biological evolution, and the nature of life itself.

Besides affecting the life sciences, AI has influenced philosophy. Many philosophers today base their accounts of mind on AI concepts. They use these to address, for instance, the notorious mind–body problem, the conundrum of free will, and the many puzzles regarding consciousness. However, these philosophical ideas are hugely controversial. And there are deep disagreements about whether any AI system could possess *real* intelligence, creativity, or life.

Last, but not least, AI has challenged the ways in which we think about humanity—and its future. Indeed, some people worry about whether we actually have a future, because they foresee AI surpassing human intelligence across the board. Although a few thinkers welcome this prospect, most dread it: what place will remain, they ask, for human dignity and responsibility?

All these issues are explored in the following chapters.

Virtual machines

'To think about AI', someone might say, 'is to think about computers'. Well, yes and no. The computers, as such, aren't the point. It's what they *do* that matters. In other words, although AI needs *physical* machines (i.e. computers), it's best thought of as using what computer scientists call *virtual* machines.

A virtual machine isn't a machine depicted in virtual reality, nor something like a simulated car engine used to train mechanics. Rather, it's the *information-processing system* that the programmer has in mind when writing a program, and that people have in mind when using it.

A word processor, for example, is thought of by its designer, and experienced by its users, as dealing directly with words and paragraphs. But the program itself usually contains neither. And a neural network (see Chapter 4) is thought of as doing information processing *in parallel*, even though it's usually implemented in a (sequential) von Neumann computer.

That's not to say that a virtual machine is just a convenient fiction, a thing merely of our imagination. Virtual machines are actual realities. They can make things happen, both inside the system and (if linked to physical devices such as cameras or robot hands) in the outside world. AI workers trying to discover what's going wrong when a program does something unexpected only rarely consider hardware faults. Usually, they're interested in the events and causal interactions in the *virtual* machinery, or software.

Programming languages, too, are virtual machines (whose instructions have to be translated into machine code before they can be run). Some are defined in terms of lower-level programming languages, so translation is required at several levels.

That's not true only of programming languages. Virtual machines in general are comprised of patterns of activity (information processing) that exist at various levels. Moreover, it's not true only of virtual machines running on computers. We'll see in Chapter 6 that *the human mind* can be understood as the virtual machine—or rather, the set of mutually interacting virtual machines, running in parallel (and developed or learned at different times)—that is implemented in the brain.

Progress in AI requires progress in defining interesting/useful virtual machines. More *physically* powerful computers (larger, faster) are all very well. They may even be necessary for certain kinds of virtual machines to be implemented. But they can't be exploited unless *informationally* powerful virtual machines can be run on them. (Similarly, progress in neuroscience requires better understanding of what *psychological* virtual machines are being implemented by the physical neurons: see Chapter 7.)

Different sorts of external-world information are used. Every AI system needs input and output devices, if only a keyboard and a screen. Often, there are also special-purpose sensors (perhaps cameras, or pressure-sensitive whiskers) and/or effectors (perhaps