



Artificial Intelligence

A Modern Approach

Fourth Edition



Stuart
Russell
Peter
Norvig

GLOBAL
EDITION

Artificial Intelligence

A Modern Approach

Fourth Edition

Global Edition



**PEARSON SERIES
IN ARTIFICIAL INTELLIGENCE**
Stuart Russell and Peter Norvig, Editors

FORSYTH & PONCE
GRAHAM
JURAFSKY & MARTIN
NEAPOLITAN
RUSSELL & NORVIG

Computer Vision: A Modern Approach, 2nd ed.
ANSI Common Lisp
Speech and Language Processing, 2nd ed.
Learning Bayesian Networks
Artificial Intelligence: A Modern Approach, 4th ed.

Artificial Intelligence

A Modern Approach

Fourth Edition
Global Edition

Stuart J. Russell and Peter Norvig

Contributing writers:

Ming-Wei Chang
Jacob Devlin
Anca Dragan
David Forsyth
Ian Goodfellow
Jitendra M. Malik
Vikash Mansinghka
Judea Pearl
Michael Wooldridge



Cover Image credits: Alan Turing: Science History Images/Alamy Stock Photo; Statue of Aristotle: Panos Karas/Shutterstock; Ada Lovelace – Pictorial Press Ltd/Alamy Stock Photo; Autonomous cars: Andrey Suslov/Shutterstock; Atlas Robot: Boston Dynamics, Inc.; Berkeley Campanile and Golden Gate Bridge: Ben Chu/Shutterstock; Background ghosted nodes: Eugene Sergeev/Alamy Stock Photo; Chess board with chess figure: Titania/Shutterstock; Mars Rover: Stocktrek Images, Inc./Alamy Stock Photo; Kasparov: KATHY WILLENS/AP Images

Pearson Education Limited

KAO Two
KAO Park
Hockham Way
Harlow
CM17 9SR
United Kingdom

and Associated Companies throughout the world

Visit us on the World Wide Web at: www.pearsonglobaleditions.com

Please contact <https://support.pearson.com/getsupport/s/contactsupport> with any queries on this content

© Pearson Education Limited 2022

The rights of Stuart Russell and Peter Norvig to be identified as the authors of this work have been asserted by them in accordance with the Copyright, Designs and Patents Act 1988.

Authorized adaptation from the United States edition, entitled Artificial Intelligence: A Modern Approach, 4th Edition, ISBN 978-0-13-461099-3 by Stuart J. Russell and Peter Norvig, published by Pearson Education © 2021.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without either the prior written permission of the publisher or a license permitting restricted copying in the United Kingdom issued by the Copyright Licensing Agency Ltd, Saffron House, 6–10 Kirby Street, London EC1N 8TS. For information regarding permissions, request forms and the appropriate contacts within the Pearson Education Global Rights & Permissions department, please visit www.pearsoned.com/permissions/.

All trademarks used herein are the property of their respective owners. The use of any trademark in this text does not vest in the author or publisher any trademark ownership rights in such trademarks, nor does the use of such trademarks imply any affiliation with or endorsement of this book by such owners.

PEARSON, ALWAYS LEARNING, and MYLAB are exclusive trademarks in the U.S. and/or other countries owned by Pearson Education, Inc. or its affiliates.

Unless otherwise indicated herein, any third-party trademarks that may appear in this work are the property of their respective owners and any references to third-party trademarks, logos or other trade dress are for demonstrative or descriptive purposes only. Such references are not intended to imply any sponsorship, endorsement, authorization, or promotion of Pearson's products by the owners of such marks, or any relationship between the owner and Pearson Education, Inc. or its affiliates, authors, licensees or distributors.

ISBN 10: 1-292-40113-3

ISBN 13: 978-1-292-40113-3

eBook ISBN 13: 978-1-292-40117-1

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

Typeset by SPi Global

eBook formatted by B2R Technologies Pvt. Ltd.

For Loy, Gordon, Lucy, George, and Isaac — S.J.R.

For Kris, Isabella, and Juliet — P.N.

This page is intentionally left blank

Preface

Artificial Intelligence (AI) is a big field, and this is a big book. We have tried to explore the full breadth of the field, which encompasses logic, probability, and continuous mathematics; perception, reasoning, learning, and action; fairness, trust, social good, and safety; and applications that range from microelectronic devices to robotic planetary explorers to online services with billions of users.

The subtitle of this book is “A Modern Approach.” That means we have chosen to tell the story from a current perspective. We synthesize what is now known into a common framework, recasting early work using the ideas and terminology that are prevalent today. We apologize to those whose subfields are, as a result, less recognizable.

New to this edition

This edition reflects the changes in AI since the last edition in 2010:

- We focus more on machine learning rather than hand-crafted knowledge engineering, due to the increased availability of data, computing resources, and new algorithms.
- Deep learning, probabilistic programming, and multiagent systems receive expanded coverage, each with their own chapter.
- The coverage of natural language understanding, robotics, and computer vision has been revised to reflect the impact of deep learning.
- The robotics chapter now includes robots that interact with humans and the application of reinforcement learning to robotics.
- Previously we defined the goal of AI as creating systems that try to maximize expected utility, where the specific utility information—the objective—is supplied by the human designers of the system. Now we no longer assume that the objective is fixed and known by the AI system; instead, the system may be uncertain about the true objectives of the humans on whose behalf it operates. It must learn what to maximize and must function appropriately even while uncertain about the objective.
- We increase coverage of the impact of AI on society, including the vital issues of ethics, fairness, trust, and safety.
- We have moved the exercises from the end of each chapter to an online site. This allows us to continuously add to, update, and improve the exercises, to meet the needs of instructors and to reflect advances in the field and in AI-related software tools.
- Overall, about 25% of the material in the book is brand new. The remaining 75% has been largely rewritten to present a more unified picture of the field. 22% of the citations in this edition are to works published after 2010.

Overview of the book

The main unifying theme is the idea of an **intelligent agent**. We define AI as the study of agents that receive percepts from the environment and perform actions. Each such agent implements a function that maps percept sequences to actions, and we cover different ways to represent these functions, such as reactive agents, real-time planners, decision-theoretic

systems, and deep learning systems. We emphasize learning both as a construction method for competent systems and as a way of extending the reach of the designer into unknown environments. We treat robotics and vision not as independently defined problems, but as occurring in the service of achieving goals. We stress the importance of the task environment in determining the appropriate agent design.

Our primary aim is to convey the *ideas* that have emerged over the past seventy years of AI research and the past two millennia of related work. We have tried to avoid excessive formality in the presentation of these ideas, while retaining precision. We have included mathematical formulas and pseudocode algorithms to make the key ideas concrete; mathematical concepts and notation are described in Appendix A and our pseudocode is described in Appendix B.

This book is primarily intended for use in an undergraduate course or course sequence. The book has 29 chapters, each requiring about a week's worth of lectures, so working through the whole book requires a two-semester sequence. A one-semester course can use selected chapters to suit the interests of the instructor and students. The book can also be used in a graduate-level course (perhaps with the addition of some of the primary sources suggested in the bibliographical notes), or for self-study or as a reference.



Term

Throughout the book, *important points* are marked with a triangle icon in the margin. Wherever a new **term** is defined, it is also noted in the margin. Subsequent significant uses of the **term** are in bold, but not in the margin. We have included a comprehensive index and an extensive bibliography.

The only prerequisite is familiarity with basic concepts of computer science (algorithms, data structures, complexity) at a sophomore level. Freshman calculus and linear algebra are useful for some of the topics.

Online resources

Online resources are available through pearsonglobaleditions.com. There you will find:

- Exercises, programming projects, and research projects. These are no longer at the end of each chapter; they are online only. Within the book, we refer to an online exercise with a name like “Exercise 6.NARY.” Instructions on the Web site allow you to find exercises by name or by topic.
- Implementations of the algorithms in the book in Python, Java, and other programming languages.
- Supplementary material and links for students and instructors.
- Instructions on how to report errors in the book in the likely event that some exist.

Book cover

The cover depicts the final position from the decisive game 6 of the 1997 chess match in which the program Deep Blue defeated Garry Kasparov (playing Black), making this the first time a computer had beaten a world champion in a chess match. Kasparov is shown at the top. To his right is a pivotal position from the second game of the historic Go match between former world champion Lee Sedol and DeepMind’s ALPHAGO program. Move 37 by ALPHAGO violated centuries of Go orthodoxy and was immediately seen by human experts

as an embarrassing mistake, but it turned out to be a winning move. At top left is an Atlas humanoid robot built by Boston Dynamics. A depiction of a self-driving car sensing its environment appears between Ada Lovelace, the world's first computer programmer, and Alan Turing, whose fundamental work defined artificial intelligence. At the bottom of the chess board are a Mars Exploration Rover robot and a statue of Aristotle, who pioneered the study of logic; his planning algorithm from *De Motu Animalium* appears behind the authors' names. Behind the chess board is a probabilistic programming model used by the UN Comprehensive Nuclear-Test-Ban Treaty Organization for detecting nuclear explosions from seismic signals.

Acknowledgments

It takes a global village to make a book. Over 600 people read parts of the book and made suggestions for improvement. The complete list is at pearsonglobaleditions.com; we are grateful to all of them. We have space here to mention only a few especially important contributors. First the contributing writers:

- Judea Pearl (Section 13.5, Causal Networks);
- Michael Wooldridge (Chapter 17, Multiagent Decision Making);
- Vikash Mansinghka (Section 18.4, Programs as Probability Models);
- Ian Goodfellow (Chapter 22, Deep Learning);
- Jacob Devlin and Mei-Wing Chang (Chapter 25, Deep Learning for Natural Language Processing);
- Anca Dragan (Chapter 26, Robotics);
- Jitendra Malik and David Forsyth (Chapter 27, Computer Vision).

Then some key roles:

- Cynthia Yeung and Malika Cantor (project management);
- Julie Sussman and Tom Galloway (copyediting and writing suggestions);
- Omari Stephens (illustrations);
- Tracy Johnson (editor);
- Erin Ault and Rose Kernan (cover and color conversion);
- Nalin Chhibber, Sam Goto, Raymond de Lacaze, Ravi Mohan, Ciaran O'Reilly, Amit Patel, Dragomir Radiv, and Samagra Sharma (online code development and mentoring);
- Google Summer of Code students (online code development).

Stuart would like to thank his wife, Loy Sheflott, for her endless patience and boundless wisdom. He hopes that Gordon, Lucy, George, and Isaac will soon be reading this book after they have forgiven him for working so long on it. RUGS (Russell's Unusual Group of Students) have been unusually helpful, as always.

Peter would like to thank his parents (Torsten and Gerda) for getting him started, and his wife (Kris), children (Bella and Juliet), colleagues, boss, and friends for encouraging and tolerating him through the long hours of writing and rewriting.

About the Authors

Stuart Russell was born in 1962 in Portsmouth, England. He received his B.A. with first-class honours in physics from Oxford University in 1982, and his Ph.D. in computer science from Stanford in 1986. He then joined the faculty of the University of California at Berkeley, where he is a professor and former chair of computer science, director of the Center for Human-Compatible AI, and holder of the Smith-Zadeh Chair in Engineering. In 1990, he received the Presidential Young Investigator Award of the National Science Foundation, and in 1995 he was cowinner of the Computers and Thought Award. He is a Fellow of the American Association for Artificial Intelligence, the Association for Computing Machinery, and the American Association for the Advancement of Science, an Honorary Fellow of Wadham College, Oxford, and an Andrew Carnegie Fellow. He held the Chaire Blaise Pascal in Paris from 2012 to 2014. He has published over 300 papers on a wide range of topics in artificial intelligence. His other books include *The Use of Knowledge in Analogy and Induction*, *Do the Right Thing: Studies in Limited Rationality* (with Eric Wefald), and *Human Compatible: Artificial Intelligence and the Problem of Control*.

Peter Norvig is currently a Director of Research at Google, Inc., and was previously the director responsible for the core Web search algorithms. He co-taught an online AI class that signed up 160,000 students, helping to kick off the current round of massive open online classes. He was head of the Computational Sciences Division at NASA Ames Research Center, overseeing research and development in artificial intelligence and robotics. He received a B.S. in applied mathematics from Brown University and a Ph.D. in computer science from Berkeley. He has been a professor at the University of Southern California and a faculty member at Berkeley and Stanford. He is a Fellow of the American Association for Artificial Intelligence, the Association for Computing Machinery, the American Academy of Arts and Sciences, and the California Academy of Science. His other books are *Paradigms of AI Programming: Case Studies in Common Lisp*, *Verbmobil: A Translation System for Face-to-Face Dialog*, and *Intelligent Help Systems for UNIX*.

The two authors shared the inaugural AAAI/EAAI Outstanding Educator award in 2016.

Contents

I Artificial Intelligence

1	Introduction	19
1.1	What Is AI?	19
1.2	The Foundations of Artificial Intelligence	23
1.3	The History of Artificial Intelligence	35
1.4	The State of the Art	45
1.5	Risks and Benefits of AI	49
	Summary	52
	Bibliographical and Historical Notes	53
2	Intelligent Agents	54
2.1	Agents and Environments	54
2.2	Good Behavior: The Concept of Rationality	57
2.3	The Nature of Environments	60
2.4	The Structure of Agents	65
	Summary	78
	Bibliographical and Historical Notes	78

II Problem-solving

3	Solving Problems by Searching	81
3.1	Problem-Solving Agents	81
3.2	Example Problems	84
3.3	Search Algorithms	89
3.4	Uninformed Search Strategies	94
3.5	Informed (Heuristic) Search Strategies	102
3.6	Heuristic Functions	115
	Summary	122
	Bibliographical and Historical Notes	124
4	Search in Complex Environments	128
4.1	Local Search and Optimization Problems	128
4.2	Local Search in Continuous Spaces	137
4.3	Search with Nondeterministic Actions	140
4.4	Search in Partially Observable Environments	144
4.5	Online Search Agents and Unknown Environments	152
	Summary	159
	Bibliographical and Historical Notes	160
5	Constraint Satisfaction Problems	164
5.1	Defining Constraint Satisfaction Problems	164
5.2	Constraint Propagation: Inference in CSPs	169

5.3	Backtracking Search for CSPs	175
5.4	Local Search for CSPs	181
5.5	The Structure of Problems	183
	Summary	187
	Bibliographical and Historical Notes	188
6	Adversarial Search and Games	192
6.1	Game Theory	192
6.2	Optimal Decisions in Games	194
6.3	Heuristic Alpha–Beta Tree Search	202
6.4	Monte Carlo Tree Search	207
6.5	Stochastic Games	210
6.6	Partially Observable Games	214
6.7	Limitations of Game Search Algorithms	219
	Summary	220
	Bibliographical and Historical Notes	221
III Knowledge, reasoning, and planning		
7	Logical Agents	226
7.1	Knowledge-Based Agents	227
7.2	The Wumpus World	228
7.3	Logic	232
7.4	Propositional Logic: A Very Simple Logic	235
7.5	Propositional Theorem Proving	240
7.6	Effective Propositional Model Checking	250
7.7	Agents Based on Propositional Logic	255
	Summary	264
	Bibliographical and Historical Notes	265
8	First-Order Logic	269
8.1	Representation Revisited	269
8.2	Syntax and Semantics of First-Order Logic	274
8.3	Using First-Order Logic	283
8.4	Knowledge Engineering in First-Order Logic	289
	Summary	295
	Bibliographical and Historical Notes	296
9	Inference in First-Order Logic	298
9.1	Propositional vs. First-Order Inference	298
9.2	Unification and First-Order Inference	300
9.3	Forward Chaining	304
9.4	Backward Chaining	311
9.5	Resolution	316
	Summary	327
	Bibliographical and Historical Notes	328

10 Knowledge Representation	332
10.1 Ontological Engineering	332
10.2 Categories and Objects	335
10.3 Events	340
10.4 Mental Objects and Modal Logic	344
10.5 Reasoning Systems for Categories	347
10.6 Reasoning with Default Information	351
Summary	355
Bibliographical and Historical Notes	356
11 Automated Planning	362
11.1 Definition of Classical Planning	362
11.2 Algorithms for Classical Planning	366
11.3 Heuristics for Planning	371
11.4 Hierarchical Planning	374
11.5 Planning and Acting in Nondeterministic Domains	383
11.6 Time, Schedules, and Resources	392
11.7 Analysis of Planning Approaches	396
Summary	397
Bibliographical and Historical Notes	398
IV Uncertain knowledge and reasoning	
12 Quantifying Uncertainty	403
12.1 Acting under Uncertainty	403
12.2 Basic Probability Notation	406
12.3 Inference Using Full Joint Distributions	413
12.4 Independence	415
12.5 Bayes' Rule and Its Use	417
12.6 Naive Bayes Models	420
12.7 The Wumpus World Revisited	422
Summary	425
Bibliographical and Historical Notes	426
13 Probabilistic Reasoning	430
13.1 Representing Knowledge in an Uncertain Domain	430
13.2 The Semantics of Bayesian Networks	432
13.3 Exact Inference in Bayesian Networks	445
13.4 Approximate Inference for Bayesian Networks	453
13.5 Causal Networks	467
Summary	471
Bibliographical and Historical Notes	472
14 Probabilistic Reasoning over Time	479
14.1 Time and Uncertainty	479
14.2 Inference in Temporal Models	483

14.3	Hidden Markov Models	491
14.4	Kalman Filters	497
14.5	Dynamic Bayesian Networks	503
	Summary	514
	Bibliographical and Historical Notes	515
15	Making Simple Decisions	518
15.1	Combining Beliefs and Desires under Uncertainty	518
15.2	The Basis of Utility Theory	519
15.3	Utility Functions	522
15.4	Multiattribute Utility Functions	530
15.5	Decision Networks	534
15.6	The Value of Information	537
15.7	Unknown Preferences	543
	Summary	547
	Bibliographical and Historical Notes	547
16	Making Complex Decisions	552
16.1	Sequential Decision Problems	552
16.2	Algorithms for MDPs	562
16.3	Bandit Problems	571
16.4	Partially Observable MDPs	578
16.5	Algorithms for Solving POMDPs	580
	Summary	585
	Bibliographical and Historical Notes	586
17	Multiagent Decision Making	589
17.1	Properties of Multiagent Environments	589
17.2	Non-Cooperative Game Theory	595
17.3	Cooperative Game Theory	616
17.4	Making Collective Decisions	622
	Summary	635
	Bibliographical and Historical Notes	636
18	Probabilistic Programming	641
18.1	Relational Probability Models	642
18.2	Open-Universe Probability Models	648
18.3	Keeping Track of a Complex World	655
18.4	Programs as Probability Models	660
	Summary	664
	Bibliographical and Historical Notes	665
V	Machine Learning	
19	Learning from Examples	669
19.1	Forms of Learning	669

19.2	Supervised Learning	671
19.3	Learning Decision Trees	675
19.4	Model Selection and Optimization	683
19.5	The Theory of Learning	690
19.6	Linear Regression and Classification	694
19.7	Nonparametric Models	704
19.8	Ensemble Learning	714
19.9	Developing Machine Learning Systems	722
	Summary	732
	Bibliographical and Historical Notes	733
20	Knowledge in Learning	739
20.1	A Logical Formulation of Learning	739
20.2	Knowledge in Learning	747
20.3	Explanation-Based Learning	750
20.4	Learning Using Relevance Information	754
20.5	Inductive Logic Programming	758
	Summary	767
	Bibliographical and Historical Notes	768
21	Learning Probabilistic Models	772
21.1	Statistical Learning	772
21.2	Learning with Complete Data	775
21.3	Learning with Hidden Variables: The EM Algorithm	788
	Summary	797
	Bibliographical and Historical Notes	798
22	Deep Learning	801
22.1	Simple Feedforward Networks	802
22.2	Computation Graphs for Deep Learning	807
22.3	Convolutional Networks	811
22.4	Learning Algorithms	816
22.5	Generalization	819
22.6	Recurrent Neural Networks	823
22.7	Unsupervised Learning and Transfer Learning	826
22.8	Applications	833
	Summary	835
	Bibliographical and Historical Notes	836
23	Reinforcement Learning	840
23.1	Learning from Rewards	840
23.2	Passive Reinforcement Learning	842
23.3	Active Reinforcement Learning	848
23.4	Generalization in Reinforcement Learning	854
23.5	Policy Search	861
23.6	Apprenticeship and Inverse Reinforcement Learning	863

23.7 Applications of Reinforcement Learning	866
Summary	869
Bibliographical and Historical Notes	870
VI Communicating, perceiving, and acting	
24 Natural Language Processing	874
24.1 Language Models	874
24.2 Grammar	884
24.3 Parsing	886
24.4 Augmented Grammars	892
24.5 Complications of Real Natural Language	896
24.6 Natural Language Tasks	900
Summary	901
Bibliographical and Historical Notes	902
25 Deep Learning for Natural Language Processing	907
25.1 Word Embeddings	907
25.2 Recurrent Neural Networks for NLP	911
25.3 Sequence-to-Sequence Models	915
25.4 The Transformer Architecture	919
25.5 Pretraining and Transfer Learning	922
25.6 State of the art	926
Summary	929
Bibliographical and Historical Notes	929
26 Robotics	932
26.1 Robots	932
26.2 Robot Hardware	933
26.3 What kind of problem is robotics solving?	937
26.4 Robotic Perception	938
26.5 Planning and Control	945
26.6 Planning Uncertain Movements	963
26.7 Reinforcement Learning in Robotics	965
26.8 Humans and Robots	968
26.9 Alternative Robotic Frameworks	975
26.10 Application Domains	978
Summary	981
Bibliographical and Historical Notes	982
27 Computer Vision	988
27.1 Introduction	988
27.2 Image Formation	989
27.3 Simple Image Features	995
27.4 Classifying Images	1002
27.5 Detecting Objects	1006

27.6	The 3D World	1008
27.7	Using Computer Vision	1013
	Summary	1026
	Bibliographical and Historical Notes	1027

VII Conclusions

28	Philosophy, Ethics, and Safety of AI	1032
28.1	The Limits of AI	1032
28.2	Can Machines Really Think?	1035
28.3	The Ethics of AI	1037
	Summary	1056
	Bibliographical and Historical Notes	1057

29	The Future of AI	1063
29.1	AI Components	1063
29.2	AI Architectures	1069

A	Mathematical Background	1074
A.1	Complexity Analysis and O() Notation	1074
A.2	Vectors, Matrices, and Linear Algebra	1076
A.3	Probability Distributions	1078
	Bibliographical and Historical Notes	1080

B	Notes on Languages and Algorithms	1081
B.1	Defining Languages with Backus–Naur Form (BNF)	1081
B.2	Describing Algorithms with Pseudocode	1082
B.3	Online Supplemental Material	1083

Bibliography	1084
---------------------	-------------

Index	1119
--------------	-------------

This page is intentionally left blank