Artificial Intelligence
A Modern Approach
Third Edition

Stuart J. Russell and Peter Norvig

Contributing writers:
Ernest Davis
Douglas D. Edwards
David Forsyth
Nicholas J. Hay
Jitendra M. Malik
Vibhu Mittal
Mehran Sahami
Sebastian Thrun
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; and everything from microelectronic devices to robotic planetary explorers. The book is also big because we go into some depth.

The subtitle of this book is “A Modern Approach.” The intended meaning of this rather empty phrase is that we have tried to synthesize what is now known into a common framework, rather than trying to explain each subfield of AI in its own historical context. We apologize to those whose subfields are, as a result, less recognizable.

New to this edition

This edition captures the changes in AI that have taken place since the last edition in 2003. There have been important applications of AI technology, such as the widespread deployment of practical speech recognition, machine translation, autonomous vehicles, and household robotics. There have been algorithmic landmarks, such as the solution of the game of checkers. And there has been a great deal of theoretical progress, particularly in areas such as probabilistic reasoning, machine learning, and computer vision. Most important from our point of view is the continued evolution in how we think about the field, and thus how we organize the book. The major changes are as follows:

- We place more emphasis on partially observable and nondeterministic environments, especially in the nonprobabilistic settings of search and planning. The concepts of belief state (a set of possible worlds) and state estimation (maintaining the belief state) are introduced in these settings; later in the book, we add probabilities.
- In addition to discussing the types of environments and types of agents, we now cover in more depth the types of representations that an agent can use. We distinguish among atomic representations (in which each state of the world is treated as a black box), factored representations (in which a state is a set of attribute/value pairs), and structured representations (in which the world consists of objects and relations between them).
- Our coverage of planning goes into more depth on contingent planning in partially observable environments and includes a new approach to hierarchical planning.
- We have added new material on first-order probabilistic models, including open-universe models for cases where there is uncertainty as to what objects exist.
- We have completely rewritten the introductory machine-learning chapter, stressing a wider variety of more modern learning algorithms and placing them on a firmer theoretical footing.
- We have expanded coverage of Web search and information extraction, and of techniques for learning from very large data sets.
- 20% of the citations in this edition are to works published after 2003.
- We estimate that about 20% of the material is brand new. The remaining 80% reflects older work but has been largely rewritten to present a more unified picture of the field.
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, and decision-theoretic systems. We explain the role of learning as extending the reach of the designer into unknown environments, and we show how that role constrains agent design, favoring explicit knowledge representation and reasoning. 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 fifty 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 pseudocode algorithms to make the key ideas concrete; our pseudocode is described in Appendix B.

This book is primarily intended for use in an undergraduate course or course sequence. The book has 27 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). Sample syllabi are available at the book’s Web site, aima.cs.berkeley.edu. 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; the required mathematical background is supplied in Appendix A.

Exercises are given at the end of each chapter. Exercises requiring significant programming are marked with a keyboard icon. These exercises can best be solved by taking advantage of the code repository at aima.cs.berkeley.edu. Some of them are large enough to be considered term projects. A number of exercises require some investigation of the literature; these are marked with a book icon.

Throughout the book, important points are marked with a pointing icon. We have included an extensive index of around 6,000 items to make it easy to find things in the book. Wherever a new term is first defined, it is also marked in the margin.

About the Web site

aima.cs.berkeley.edu, the Web site for the book, contains

- implementations of the algorithms in the book in several programming languages,
- a list of over 1000 schools that have used the book, many with links to online course materials and syllabi,
- an annotated list of over 800 links to sites around the Web with useful AI content,
- a chapter-by-chapter list of supplementary material and links,
- instructions on how to join a discussion group for the book,
• instructions on how to contact the authors with questions or comments,
• instructions on how to report errors in the book, in the likely event that some exist, and
• slides and other materials for instructors.

Pearson offers many different products around the world to facilitate learning. In countries outside the United States, some products and services related to this textbook may not be available due to copyright and/or permissions restrictions. If you have questions, you can contact your local office by visiting www.pearsonhighered.com/international or you can contact your local Pearson representative.

About the cover

The cover depicts the final position from the decisive game 6 of the 1997 match between chess champion Garry Kasparov and program DEEP BLUE. Kasparov, playing Black, was forced to resign, making this the first time a computer had beaten a world champion in a chess match. Kasparov is shown at the top. To his left is the Asimo humanoid robot and to his right is Thomas Bayes (1702–1761), whose ideas about probability as a measure of belief underlie much of modern AI technology. Below that we see a Mars Exploration Rover, a robot that landed on Mars in 2004 and has been exploring the planet ever since. To the right is Alan Turing (1912–1954), whose fundamental work defined the fields of computer science in general and artificial intelligence in particular. At the bottom is Shakey (1966–1972), the first robot to combine perception, world-modeling, planning, and learning. With Shakey is project leader Charles Rosen (1917–2002). At the bottom right is Aristotle (384 B.C.–322 B.C.), who pioneered the study of logic; his work was state of the art until the 19th century (copy of a bust by Lysippos). At the bottom left, lightly screened behind the authors’ names, is a planning algorithm by Aristotle from De Motu Animalium in the original Greek. Behind the title is a portion of the CPSC Bayesian network for medical diagnosis (Pradhan et al., 1994). Behind the chess board is part of a Bayesian logic model for detecting nuclear explosions from seismic signals.

Credits: Stan Honda/ Getty (Kasparaov), Library of Congress (Bayes), NASA (Mars rover), National Museum of Rome (Aristotle), Peter Norvig (book), Ian Parker (Berkeley skyline), Shutterstock (Asimo, Chess pieces), Time Life/ Getty (Shakey, Turing).

Acknowledgments

This book would not have been possible without the many contributors whose names did not make it to the cover. Jitendra Malik and David Forsyth wrote Chapter 24 (computer vision) and Sebastian Thrun wrote Chapter 25 (robotics). Vibhu Mittal wrote part of Chapter 22 (natural language). Nick Hay, Mehran Sahami, and Ernest Davis wrote some of the exercises. Zoran Duric (George Mason), Thomas C. Henderson (Utah), Leon Reznik (RIT), Michael Gourley (Central Oklahoma) and Ernest Davis (NYU) reviewed the manuscript and made helpful suggestions. We thank Ernie Davis in particular for his tireless ability to read multiple drafts and help improve the book. Nick Hay whipped the bibliography into shape and on deadline stayed up to 5:30 AM writing code to make the book better. Jon Barron formatted and improved the diagrams in this edition, while Tim Huang, Mark Paskin, and Cynthia
Bruyns helped with diagrams and algorithms in previous editions. Ravi Mohan and Ciaran O’Reilly wrote and maintain the Java code examples on the Web site. John Canny wrote the robotics chapter for the first edition and Douglas Edwards researched the historical notes. Tracy Dunkelberger, Allison Michael, Scott Disanno, and Jane Bonnell at Pearson tried their best to keep us on schedule and made many helpful suggestions. Most helpful of all has been Julie Sussman, P.P.A., who read every chapter and provided extensive improvements. In previous editions we had proofreaders who would tell us when we left out a comma and said which when we meant that; Julie told us when we left out a minus sign and said $x_i$ when we meant $x_j$. For every typo or confusing explanation that remains in the book, rest assured that Julie has fixed at least five. She persevered even when a power failure forced her to work by lantern light rather than LCD glow.

**Stuart would like to thank** his parents for their support and encouragement and 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, and friends for encouraging and tolerating him through the long hours of writing and longer hours of 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 of computer science, director of the Center for Intelligent Systems, 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 was a 1996 Miller Professor of the University of California and was appointed to a Chancellor’s Professorship in 2000. In 1998, he gave the Forsythe Memorial Lectures at Stanford University. He is a Fellow and former Executive Council member of the American Association for Artificial Intelligence. He has published over 100 papers on a wide range of topics in artificial intelligence. His other books include *The Use of Knowledge in Analogy and Induction* and (with Eric Wefald) *Do the Right Thing: Studies in Limited Rationality.*

**Peter Norvig** is currently Director of Research at Google, Inc., and was the director responsible for the core Web search algorithms from 2002 to 2005. He is a Fellow of the American Association for Artificial Intelligence and the Association for Computing Machinery. Previously, he was head of the Computational Sciences Division at NASA Ames Research Center, where he oversaw NASA’s research and development in artificial intelligence and robotics, and chief scientist at Junglee, where he helped develop one of the first Internet information extraction services. He received a B.S. in applied mathematics from Brown University and a Ph.D. in computer science from the University of California at Berkeley. He received the Distinguished Alumni and Engineering Innovation awards from Berkeley and the Exceptional Achievement Medal from NASA. He has been a professor at the University of Southern California and a research faculty member at Berkeley. His other books are *Paradigms of AI Programming: Case Studies in Common Lisp* and *Verbmobil: A Translation System for Face-to-Face Dialog* and *Intelligent Help Systems for UNIX.*
## Contents

### I Artificial Intelligence

1 Introduction

1.1 What Is AI? ......................................................... 1
1.2 The Foundations of Artificial Intelligence ................. 5
1.3 The History of Artificial Intelligence ...................... 16
1.4 The State of the Art ........................................... 28
1.5 Summary, Bibliographical and Historical Notes, Exercises 29

2 Intelligent Agents ................................................. 34

2.1 Agents and Environments .................................. 34
2.2 Good Behavior: The Concept of Rationality ............. 36
2.3 The Nature of Environments ................................. 40
2.4 The Structure of Agents .................................. 46
2.5 Summary, Bibliographical and Historical Notes, Exercises 59

### II Problem-solving

3 Solving Problems by Searching .................................. 64

3.1 Problem-Solving Agents .................................. 64
3.2 Example Problems ......................................... 69
3.3 Searching for Solutions ................................... 75
3.4 Uninformed Search Strategies ............................ 81
3.5 Informed (Heuristic) Search Strategies .................. 92
3.6 Heuristic Functions ....................................... 102
3.7 Summary, Bibliographical and Historical Notes, Exercises 108

4 Beyond Classical Search ........................................ 120

4.1 Local Search Algorithms and Optimization Problems .... 120
4.2 Local Search in Continuous Spaces ....................... 129
4.3 Searching with Nondeterministic Actions ............... 133
4.4 Searching with Partial Observations ..................... 138
4.5 Online Search Agents and Unknown Environments .... 147
4.6 Summary, Bibliographical and Historical Notes, Exercises 153

5 Adversarial Search ................................................. 161

5.1 Games ......................................................... 161
5.2 Optimal Decisions in Games ............................... 163
5.3 Alpha–Beta Pruning ........................................ 167
5.4 Imperfect Real-Time Decisions ............................ 171
5.5 Stochastic Games ......................................... 177
10.4 Other Classical Planning Approaches .................................................. 387
10.5 Analysis of Planning Approaches ...................................................... 392
10.6 Summary, Bibliographical and Historical Notes, Exercises .................. 393

11 Planning and Acting in the Real World .............................................. 401
11.1 Time, Schedules, and Resources ....................................................... 401
11.2 Hierarchical Planning ................................................................. 406
11.3 Planning and Acting in Nondeterministic Domains ............................. 415
11.4 Multiagent Planning ................................................................. 425
11.5 Summary, Bibliographical and Historical Notes, Exercises .................. 430

12 Knowledge Representation ............................................................... 437
12.1 Ontological Engineering ............................................................... 437
12.2 Categories and Objects ............................................................... 440
12.3 Events ......................................................................................... 446
12.4 Mental Events and Mental Objects .................................................. 450
12.5 Reasoning Systems for Categories .................................................. 453
12.6 Reasoning with Default Information ............................................... 458
12.7 The Internet Shopping World ........................................................ 462
12.8 Summary, Bibliographical and Historical Notes, Exercises .................. 467

IV Uncertain knowledge and reasoning .................................................. 480
13 Quantifying Uncertainty ................................................................. 480
13.1 Acting under Uncertainty ............................................................... 480
13.2 Basic Probability Notation ............................................................. 483
13.3 Inference Using Full Joint Distributions ........................................... 490
13.4 Independence .............................................................................. 494
13.5 Bayes’ Rule and Its Use ............................................................... 495
13.6 The Wumpus World Revisited ....................................................... 499
13.7 Summary, Bibliographical and Historical Notes, Exercises .............. 503

14 Probabilistic Reasoning ................................................................. 510
14.1 Representing Knowledge in an Uncertain Domain ............................. 510
14.2 The Semantics of Bayesian Networks ............................................. 513
14.3 Efficient Representation of Conditional Distributions ....................... 518
14.4 Exact Inference in Bayesian Networks ............................................. 522
14.5 Approximate Inference in Bayesian Networks .................................. 530
14.6 Relational and First-Order Probability Models .................................. 539
14.7 Other Approaches to Uncertain Reasoning ..................................... 546
14.8 Summary, Bibliographical and Historical Notes, Exercises .............. 551

15 Probabilistic Reasoning over Time ................................................... 566
15.1 Time and Uncertainty ................................................................. 566
15.2 Inference in Temporal Models .............................................. 570
15.3 Hidden Markov Models ..................................................... 578
15.4 Kalman Filters ............................................................... 584
15.5 Dynamic Bayesian Networks ............................................... 590
15.6 Keeping Track of Many Objects .......................................... 599
15.7 Summary, Bibliographical and Historical Notes, Exercises ............ 603

16 Making Simple Decisions .................................................... 610
16.1 Combining Beliefs and Desires under Uncertainty ....................... 610
16.2 The Basis of Utility Theory ............................................... 611
16.3 Utility Functions ............................................................. 615
16.4 Multiattribute Utility Functions ............................................ 622
16.5 Decision Networks .......................................................... 626
16.6 The Value of Information .................................................. 628
16.7 Decision-Theoretic Expert Systems ....................................... 633
16.8 Summary, Bibliographical and Historical Notes, Exercises ............ 636

17 Making Complex Decisions .................................................. 645
17.1 Sequential Decision Problems ............................................. 645
17.2 Value Iteration ............................................................... 652
17.3 Policy Iteration .............................................................. 656
17.4 Partially Observable MDPs ................................................... 658
17.5 Decisions with Multiple Agents: Game Theory ............................ 666
17.6 Mechanism Design .......................................................... 679
17.7 Summary, Bibliographical and Historical Notes, Exercises ............ 684

V Learning

18 Learning from Examples ..................................................... 693
18.1 Forms of Learning ........................................................ 693
18.2 Supervised Learning ......................................................... 695
18.3 Learning Decision Trees .................................................... 697
18.4 Evaluating and Choosing the Best Hypothesis ............................ 708
18.5 The Theory of Learning ..................................................... 713
18.6 Regression and Classification with Linear Models ......................... 717
18.7 Artificial Neural Networks ................................................ 727
18.8 Nonparametric Models ....................................................... 737
18.9 Support Vector Machines ................................................... 744
18.10 Ensemble Learning .......................................................... 748
18.11 Practical Machine Learning ............................................... 753
18.12 Summary, Bibliographical and Historical Notes, Exercises ............ 757

19 Knowledge in Learning ...................................................... 768
19.1 A Logical Formulation of Learning ....................................... 768
## 19 Knowledge in Learning
- 19.2 Knowledge in Learning .................................................. 777
- 19.3 Explanation-Based Learning ............................................. 780
- 19.4 Learning Using Relevance Information ............................... 784
- 19.5 Inductive Logic Programming ........................................... 788
- 19.6 Summary, Bibliographical and Historical Notes, Exercises ...... 797

## 20 Learning Probabilistic Models 802
- 20.1 Statistical Learning ....................................................... 802
- 20.2 Learning with Complete Data .......................................... 806
- 20.3 Learning with Hidden Variables: The EM Algorithm ............. 816
- 20.4 Summary, Bibliographical and Historical Notes, Exercises ...... 825

## 21 Reinforcement Learning 830
- 21.1 Introduction ............................................................... 830
- 21.2 Passive Reinforcement Learning ....................................... 832
- 21.3 Active Reinforcement Learning ........................................ 839
- 21.4 Generalization in Reinforcement Learning ......................... 845
- 21.5 Policy Search ............................................................. 848
- 21.6 Applications of Reinforcement Learning ............................ 850
- 21.7 Summary, Bibliographical and Historical Notes, Exercises ...... 853

## VI Communicating, perceiving, and acting

### 22 Natural Language Processing 860
- 22.1 Language Models .......................................................... 860
- 22.2 Text Classification ........................................................ 865
- 22.3 Information Retrieval ...................................................... 867
- 22.4 Information Extraction ................................................... 873
- 22.5 Summary, Bibliographical and Historical Notes, Exercises ...... 882

### 23 Natural Language for Communication 888
- 23.1 Phrase Structure Grammars ............................................. 888
- 23.2 Syntactic Analysis (Parsing) .............................................. 892
- 23.3 Augmented Grammars and Semantic Interpretation ............... 897
- 23.4 Machine Translation ...................................................... 907
- 23.5 Speech Recognition ....................................................... 912
- 23.6 Summary, Bibliographical and Historical Notes, Exercises ...... 918

### 24 Perception 928
- 24.1 Image Formation ........................................................... 929
- 24.2 Early Image-Processing Operations .................................... 935
- 24.3 Object Recognition by Appearance ..................................... 942
- 24.4 Reconstructing the 3D World .......................................... 947
- 24.5 Object Recognition from Structural Information .................. 957
25 Robotics  
25.1 Introduction 971  
25.2 Robot Hardware 973  
25.3 Robotic Perception 978  
25.4 Planning to Move 986  
25.5 Planning Uncertain Movements 993  
25.6 Moving 997  
25.7 Robotic Software Architectures 1003  
25.8 Application Domains 1006  
25.9 Summary, Bibliographical and Historical Notes, Exercises 1010

VII Conclusions

26 Philosophical Foundations 1020  
26.1 Weak AI: Can Machines Act Intelligently? 1020  
26.2 Strong AI: Can Machines Really Think? 1026  
26.3 The Ethics and Risks of Developing Artificial Intelligence 1034  
26.4 Summary, Bibliographical and Historical Notes, Exercises 1040

27 AI: The Present and Future 1044  
27.1 Agent Components 1044  
27.2 Agent Architectures 1047  
27.3 Are We Going in the Right Direction? 1049  
27.4 What If AI Does Succeed? 1051

A Mathematical background 1053  
A.1 Complexity Analysis and O() Notation 1053  
A.2 Vectors, Matrices, and Linear Algebra 1055  
A.3 Probability Distributions 1057

B Notes on Languages and Algorithms 1060  
B.1 Defining Languages with Backus–Naur Form (BNF) 1060  
B.2 Describing Algorithms with Pseudocode 1061  
B.3 Online Help 1062

Bibliography 1063

Index 1095