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 Analogical and Inductive Inference 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 JPL, 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 Verbosity: A Translation System for Face-to-Face Dialog and Intelligent Help Systems for UNIX.

Contents

1 Artificial Intelligence

1 Introduction ............................................. 1
  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

III Problem-solving ..................................... 64

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
| Contents |
|-----------------|--------|
| 5.6 Partially Observable Games | 180 |
| 5.7 State-of-the-Art Game Programs | 185 |
| 5.8 Alternative Approaches | 187 |
| 5.9 Summary, Bibliographical and Historical Notes, Exercises | 189 |
| **6 Constraint Satisfaction Problems** | **202** |
| 6.1 Defining Constraint Satisfaction Problems | 202 |
| 6.2 Constraint Propagation: Inference in CSPs | 208 |
| 6.3 Backtracking Search for CSPs | 214 |
| 6.4 Local Search for CSPs | 220 |
| 6.5 The Structure of Problems | 222 |
| 6.6 Summary, Bibliographical and Historical Notes, Exercises | 227 |
| **III Knowledge, Reasoning, and Planning** | **234** |
| 7 Logical Agents | |
| 7.1 Knowledge-Based Agents | 255 |
| 7.2 The Wumpus World | 236 |
| 7.3 Logic | 240 |
| 7.4 Propositional Logic: A Very Simple Logic | 243 |
| 7.5 Propositional Theorem Proving | 249 |
| 7.6 Effective Propositional Model Checking | 259 |
| 7.7 Agents Based on Propositional Logic | 265 |
| 7.8 Summary, Bibliographical and Historical Notes, Exercises | 274 |
| **8 First-Order Logic** | **285** |
| 8.1 Representation Revisited | 285 |
| 8.2 Syntax and Semantics of First-Order Logic | 290 |
| 8.3 Using First-Order Logic | 300 |
| 8.4 Knowledge Engineering in First-Order Logic | 307 |
| 8.5 Summary, Bibliographical and Historical Notes, Exercises | 313 |
| **9 Inference in First-Order Logic** | **322** |
| 9.1 Propositional vs. First-Order Inference | 322 |
| 9.2 Unification and Lifting | 325 |
| 9.3 Forward Chaining | 330 |
| 9.4 Backward Chaining | 337 |
| 9.5 Resolution | 345 |
| 9.6 Summary, Bibliographical and Historical Notes, Exercises | 357 |
| **10 Classical Planning** | **366** |
| 10.1 Definition of Classical Planning | 366 |
| 10.2 Algorithms for Planning as State-Space Search | 373 |
| 10.3 Planning Graphs | 379 |
| **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 Multistage 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** | |
| **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 |

*Contents*
xv

5.6 Partially Observable Games ........................................... 180
5.7 State-of-the-Art Game Programs ................................. 185
5.8 Alternative Approaches ............................................... 187
5.9 Summary, Bibliographical and Historical Notes, Exercises ........... 189

6 Constraint Satisfaction Problems ................................. 202
6.1 Defining Constraint Satisfaction Problems ....................... 202
6.2 Constraint Propagation: Inference in CSPs ...................... 208
6.3 Backtracking Search for CSPs .................................... 214
6.4 Local Search for CSPs ........................................... 220
6.5 The Structure of Problems .................................... 222
6.6 Summary, Bibliographical and Historical Notes, Exercises .......... 227

III Knowledge, reasoning, and planning

7 Logical Agents ........................................................... 234
7.1 Knowledge-Based Agents ............................................ 235
7.2 The Wumpus World ................................................ 236
7.3 Logic .................................................................... 240
7.4 Propositional Logic: A Very Simple Logic .................. 243
7.5 Propositional Theorem Proving ................................ 249
7.6 Effective Propositional Model Checking ..................... 259
7.7 Agents Based on Propositional Logic ............................ 265
7.8 Summary, Bibliographical and Historical Notes, Exercises .......... 274

8 First-Order Logic .......................................................... 285
8.1 Representation Revisited ........................................... 285
8.2 Syntax and Semantics of First-Order Logic .................... 290
8.3 Using First-Order Logic .......................................... 300
8.4 Knowledge Engineering in First-Order Logic ................. 307
8.5 Summary, Bibliographical and Historical Notes, Exercises .......... 313

9 Inference in First-Order Logic ........................................ 322
9.1 Propositional vs. First-Order Inference ........................ 322
9.2 Unification and Lifting ........................................... 325
9.3 Forward Chaining ................................................ 330
9.4 Backward Chaining ............................................... 337
9.5 Resolution .................................................................. 345
9.6 Summary, Bibliographical and Historical Notes, Exercises .......... 357

10 Classical Planning ......................................................... 366
10.1 Definition of Classical Planning ............................... 366
10.2 Algorithms for Planning as State-Space Search .......... 373
10.3 Planning Graphs ...................................................... 379

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

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 Multistate 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

19.2 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
INTRODUCTION

In which we try to explain why we consider artificial intelligence to be a subject most worthy of study, and in which we try to decide what exactly it is, this being a good thing to decide before embarking.

We call ourselves Homo sapiens—man the wise—because our intelligence is so important to us. For thousands of years, we have tried to understand how we think; that is, how a mere handful of matter can perceive, understand, predict, and manipulate a world far larger and more complicated than itself. The field of artificial intelligence, or AI, goes further still: it attempts not just to understand but also to build intelligent entities.

AI is one of the newest fields in science and engineering. Work started in earnest soon after World War II, and the name itself was coined in 1955. Along with molecular biology, AI is regularly cited as the “field I would most like to be in” by scientists in other disciplines. A student in physics might reasonably feel that all the good ideas have already been taken by Galileo, Newton, Einstein, and the rest. AI, on the other hand, still has openings for several full-time Einsteins and Einsteins.

AI currently encompasses a huge variety of subfields, ranging from the general (learning and perception) to the specific, such as playing chess, proving mathematical theorems, writing poetry, driving a car on a crowded street, and diagnosing diseases. AI is relevant to any intellectual task; it is truly a universal field.

1.1 WHAT IS AI?

We have claimed that AI is exciting, but we have not said what it is. In Figure 1.1 we see eight definitions of AI, laid out along two dimensions. The definitions on top are concerned with thought processes and reasoning, whereas the ones on the bottom address behavior. The definitions on the left measure success in terms of fidelity to human performance, whereas the ones on the right measure against an ideal performance measure, called rationality. A system is rational if it does the “right thing,” given what it knows. Historically, all four approaches to AI have been followed, each by different people with different methods. A human-centered approach must be in part an empirical science, i-