Contents

I Artificial Intelligence

1 Introduction **1**

1.1 What is AI? . . . 1
Acting humanly: The Turing Test approach . . . 2
Thinking humanly: The cognitive modeling approach . . . 3
Thinking rationally: The "laws of thought" approach . . . 4
Acting rationally: The rational agent approach . . . 4

1.2 The Foundations of Artificial Intelligence . . . 5
Philosophy (428 B.C.–present) . . . 5
Mathematics (c. 800–present) . . . 7
Economics (1776–present) . . . 9
Neuroscience (1861–present) . . . 10
Psychology (1879–present) . . . 12
Computer engineering (1940–present) . . . 14
Control theory and Cybernetics (1948–present) . . . 15
Linguistics (1957–present) . . . 16

1.3 The History of Artificial Intelligence . . . 16
The gestation of artificial intelligence (1943–1955) . . . 16
The birth of artificial intelligence (1956) . . . 17
Early enthusiasm, great expectations (1952–1969) . . . 18
A dose of reality (1966–1973) . . . 21
Knowledge-based systems: The key to power? (1969–1979) . . . 22
AI becomes an industry (1980–present) . . . 24
The return of neural networks (1986–present) . . . 25
AI becomes a science (1987–present) . . . 25
The emergence of intelligent agents (1995–present) . . . 27

1.4 The State of the Art . . . 27

1.5 Summary . . . 28

Bibliographical and Historical Notes . . . 29

Exercises . . . 30

2 Intelligent Agents **32**

2.1 Agents and Environments . . . 32

2.2 Good Behavior: The Concept of Rationality . . . 34
Performance measures . . . 35
Rationality . . . 35
Omniscience, learning, and autonomy . . . 36

2.3 The Nature of Environments . . . 38
Specifying the task environment . . . 38
Properties of task environments . . . 40

2.4 The Structure of Agents . . . 44
Agent programs . . . 44
Simple reflex agents . . . 46
Model-based reflex agents . . . 48

Goal-based agents . . . 49
Utility-based agents . . . 51
Learning agents . . . 51
2.5 Summary . . . 54
Bibliographical and Historical Notes . . . 55
Exercises . . . 56

II Problem-solving

3 Solving Problems by Searching **59**
3.1 Problem-Solving Agents . . . 59
Well-defined problems and solutions . . . 62
Formulating problems . . . 62
3.2 Example Problems . . . 64
Toy problems . . . 64
Real-world problems . . . 67
3.3 Searching for Solutions . . . 69
Measuring problem-solving performance . . . 71
3.4 Uninformed Search Strategies . . . 73
Breadth-first search . . . 73
Depth-first search . . . 75
Depth-limited search . . . 77
Iterative deepening depth-first search . . . 78
Bidirectional search . . . 79
Comparing uninformed search strategies . . . 81
3.5 Avoiding Repeated States . . . 81
3.6 Searching with Partial Information . . . 83
Sensorless problems . . . 84
Contingency problems . . . 86
3.7 Summary . . . 87
Bibliographical and Historical Notes . . . 88
Exercises . . . 89

4 Informed Search and Exploration **94**
4.1 Informed (Heuristic) Search Strategies . . . 94
Greedy best-first search . . . 95
A* search: Minimizing the total estimated solution cost . . . 97
Memory-bounded heuristic search . . . 101
Learning to search better . . . 104
4.2 Heuristic Functions . . . 105
The effect of heuristic accuracy on performance . . . 106
Inventing admissible heuristic functions . . . 107
Learning heuristics from experience . . . 109
4.3 Local Search Algorithms and Optimization Problems . . . 110
Hill-climbing search . . . 111
Simulated annealing search . . . 115
Local beam search . . . 115
Genetic algorithms . . . 116
4.4 Local Search in Continuous Spaces . . . 119

4.5 Online Search Agents and Unknown Environments . . . 122
Online search problems . . . 123
Online search agents . . . 125
Online local search . . . 126
Learning in online search . . . 127
4.6 Summary . . . 129
Bibliographical and Historical Notes . . . 130
Exercises . . . 134

5 Constraint Satisfaction Problems **137**
5.1 Constraint Satisfaction Problems . . . 137
5.2 Backtracking Search for CSPs . . . 141
Variable and value ordering . . . 143
Propagating information through constraints . . . 144
Intelligent backtracking: looking backward . . . 148
5.3 Local Search for Constraint Satisfaction Problems . . . 150
5.4 The Structure of Problems . . . 151
5.5 Summary . . . 155
Bibliographical and Historical Notes . . . 156
Exercises . . . 158

6 Adversarial Search **161**
6.1 Games . . . 161
6.2 Optimal Decisions in Games . . . 162
Optimal strategies . . . 163
The minimax algorithm . . . 165
Optimal decisions in multiplayer games . . . 165
6.3 Alpha–Beta Pruning . . . 167
6.4 Imperfect, Real-Time Decisions . . . 171
Evaluation functions . . . 171
Cutting off search . . . 173
6.5 Games That Include an Element of Chance . . . 175
Position evaluation in games with chance nodes . . . 177
Complexity of expectiminimax . . . 177
Card games . . . 179
6.6 State-of-the-Art Game Programs . . . 180
6.7 Discussion . . . 183
6.8 Summary . . . 185
Bibliographical and Historical Notes . . . 186
Exercises . . . 189

III Knowledge and reasoning

7 Logical Agents **194**
7.1 Knowledge-Based Agents . . . 195
7.2 The Wumpus World . . . 197
7.3 Logic . . . 200
7.4 Propositional Logic: A Very Simple Logic . . . 204
Syntax . . . 204

Semantics . . . 206
A simple knowledge base . . . 208
Inference . . . 208
Equivalence, validity, and satisfiability . . . 210
7.5 Reasoning Patterns in Propositional Logic . . . 211
Resolution . . . 213
Forward and backward chaining . . . 217
7.6 Effective propositional inference . . . 220
A complete backtracking algorithm . . . 221
Local-search algorithms . . . 222
Hard satisfiability problems . . . 224
7.7 Agents Based on Propositional Logic . . . 225
Finding pits and wumpuses using logical inference . . . 225
Keeping track of location and orientation . . . 227
Circuit-based agents . . . 227
A comparison . . . 231
7.8 Summary . . . 232
Bibliographical and Historical Notes . . . 233
Exercises . . . 236

8 First-Order Logic **240**
8.1 Representation Revisited . . . 240
8.2 Syntax and Semantics of First-Order Logic . . . 245
Models for first-order logic . . . 245
Symbols and interpretations . . . 246
Terms . . . 248
Atomic sentences . . . 248
Complex sentences . . . 249
Quantifiers . . . 249
Equality . . . 253
8.3 Using First-Order Logic . . . 253
Assertions and queries in first-order logic . . . 253
The kinship domain . . . 254
Numbers, sets, and lists . . . 256
The wumpus world . . . 258
8.4 Knowledge Engineering in First-Order Logic . . . 260
The knowledge engineering process . . . 261
The electronic circuits domain . . . 262
8.5 Summary . . . 266
Bibliographical and Historical Notes . . . 267
Exercises . . . 268

9 Inference in First-Order Logic **272**
9.1 Propositional vs. First-Order Inference . . . 272
Inference rules for quantifiers . . . 273
Reduction to propositional inference . . . 274
9.2 Unification and Lifting . . . 275
A first-order inference rule . . . 275
Unification . . . 276

Storage and retrieval . . . 278
9.3 Forward Chaining . . . 280
First-order definite clauses . . . 280
A simple forward-chaining algorithm . . . 281
Efficient forward chaining . . . 283
9.4 Backward Chaining . . . 287
A backward chaining algorithm . . . 287
Logic programming . . . 289
Efficient implementation of logic programs . . . 290
Redundant inference and infinite loops . . . 292
Constraint logic programming . . . 294
9.5 Resolution . . . 295
Conjunctive normal form for first-order logic . . . 295
The resolution inference rule . . . 297
Example proofs . . . 297
Completeness of resolution . . . 300
Dealing with equality . . . 303
Resolution strategies . . . 304
Theorem provers . . . 306
9.6 Summary . . . 310
Bibliographical and Historical Notes . . . 310
Exercises . . . 315

10 Knowledge Representation **320**
10.1 Ontological Engineering . . . 320
10.2 Categories and Objects . . . 322
Physical composition . . . 324
Measurements . . . 325
Substances and objects . . . 327
10.3 Actions, Situations, and Events . . . 328
The ontology of situation calculus . . . 329
Describing actions in situation calculus . . . 330
Solving the representational frame problem . . . 332
Solving the inferential frame problem . . . 333
Time and event calculus . . . 334
Generalized events . . . 335
Processes . . . 337
Intervals . . . 338
Fluents and objects . . . 339
10.4 Mental Events and Mental Objects . . . 341
A formal theory of beliefs . . . 341
Knowledge and belief . . . 343
Knowledge, time, and action . . . 344
10.5 The Internet Shopping World . . . 344
Comparing offers . . . 348
10.6 Reasoning Systems for Categories . . . 349
Semantic networks . . . 350
Description logics . . . 353
10.7 Reasoning with Default Information . . . 354

Open and closed worlds . . . 354
Negation as failure and stable model semantics . . . 356
Circumscription and default logic . . . 358
10.8 Truth Maintenance Systems . . . 360
10.9 Summary . . . 362
Bibliographical and Historical Notes . . . 363
Exercises . . . 369

IV Planning

11 Planning **375**
11.1 The Planning Problem . . . 375
The language of planning problems . . . 377
Expressiveness and extensions . . . 378
Example: Air cargo transport . . . 380
Example: The spare tire problem . . . 381
Example: The blocks world . . . 381
11.2 Planning with State-Space Search . . . 382
Forward state-space search . . . 382
Backward state-space search . . . 384
Heuristics for state-space search . . . 386
11.3 Partial-Order Planning . . . 387
A partial-order planning example . . . 391
Partial-order planning with unbound variables . . . 393
Heuristics for partial-order planning . . . 394
11.4 Planning Graphs . . . 395
Planning graphs for heuristic estimation . . . 397
The GRAPHPLAN algorithm . . . 398
Termination of GRAPHPLAN . . . 401
11.5 Planning with Propositional Logic . . . 402
Describing planning problems in propositional logic . . . 402
Complexity of propositional encodings . . . 405
11.6 Analysis of Planning Approaches . . . 407
11.7 Summary . . . 408
Bibliographical and Historical Notes . . . 409
Exercises . . . 412

12 Planning and Acting in the Real World **417**
12.1 Time, Schedules, and Resources . . . 417
Scheduling with resource constraints . . . 420
12.2 Hierarchical Task Network Planning . . . 422
Representing action decompositions . . . 423
Modifying the planner for decompositions . . . 425
Discussion . . . 427
12.3 Planning and Acting in Nondeterministic Domains . . . 430
12.4 Conditional Planning . . . 433
Conditional planning in fully observable environments . . . 433
Conditional planning in partially observable environments . . . 437
12.5 Execution Monitoring and Replanning . . . 441

12.6 Continuous Planning . . . 445
12.7 MultiAgent Planning . . . 449
Cooperation: Joint goals and plans . . . 450
Multibody planning . . . 451
Coordination mechanisms . . . 452
Competition . . . 454
12.8 Summary . . . 454
Bibliographical and Historical Notes . . . 455
Exercises . . . 459

V Uncertain knowledge and reasoning

13 Uncertainty **462**
13.1 Acting under Uncertainty . . . 462
Handling uncertain knowledge . . . 463
Uncertainty and rational decisions . . . 465
Design for a decision-theoretic agent . . . 466
13.2 Basic Probability Notation . . . 466
Propositions . . . 467
Atomic events . . . 468
Prior probability . . . 468
Conditional probability . . . 470
13.3 The Axioms of Probability . . . 471
Using the axioms of probability . . . 473
Why the axioms of probability are reasonable . . . 473
13.4 Inference Using Full Joint Distributions . . . 475
13.5 Independence . . . 477
13.6 Bayes' Rule and Its Use . . . 479
Applying Bayes' rule: The simple case . . . 480
Using Bayes' rule: Combining evidence . . . 481
13.7 The Wumpus World Revisited . . . 483
13.8 Summary . . . 486
Bibliographical and Historical Notes . . . 487
Exercises . . . 489

14 Probabilistic Reasoning **492**
14.1 Representing Knowledge in an Uncertain Domain . . . 492
14.2 The Semantics of Bayesian Networks . . . 495
Representing the full joint distribution . . . 495
Conditional independence relations in Bayesian networks . . . 499
14.3 Efficient Representation of Conditional Distributions . . . 500
14.4 Exact Inference in Bayesian Networks . . . 504
Inference by enumeration . . . 504
The variable elimination algorithm . . . 507
The complexity of exact inference . . . 509
Clustering algorithms . . . 510
14.5 Approximate Inference in Bayesian Networks . . . 511
Direct sampling methods . . . 511
Inference by Markov chain simulation . . . 516

14.6 Extending Probability to First-Order Representations . . . 519
14.7 Other Approaches to Uncertain Reasoning . . . 523
Rule-based methods for uncertain reasoning . . . 524
Representing ignorance: Dempster–Shafer theory . . . 525
Representing vagueness: Fuzzy sets and fuzzy logic . . . 526
14.8 Summary . . . 528
Bibliographical and Historical Notes . . . 528
Exercises . . . 533

15 Probabilistic Reasoning over Time **537**
15.1 Time and Uncertainty . . . 537
States and observations . . . 538
Stationary processes and the Markov assumption . . . 538
15.2 Inference in Temporal Models . . . 541
Filtering and prediction . . . 542
Smoothing . . . 544
Finding the most likely sequence . . . 547
15.3 Hidden Markov Models . . . 549
Simplified matrix algorithms . . . 549
15.4 Kalman Filters . . . 551
Updating Gaussian distributions . . . 553
A simple one-dimensional example . . . 554
The general case . . . 556
Applicability of Kalman filtering . . . 557
15.5 Dynamic Bayesian Networks . . . 559
Constructing DBNs . . . 560
Exact inference in DBNs . . . 563
Approximate inference in DBNs . . . 565
15.6 Speech Recognition . . . 568
Speech sounds . . . 570
Words . . . 572
Sentences . . . 574
Building a speech recognizer . . . 576
15.7 Summary . . . 578
Bibliographical and Historical Notes . . . 578
Exercises . . . 581

16 Making Simple Decisions **584**
16.1 Combining Beliefs and Desires under Uncertainty . . . 584
16.2 The Basis of Utility Theory . . . 586
Constraints on rational preferences . . . 586
And then there was Utility . . . 588
16.3 Utility Functions . . . 589
The utility of money . . . 589
Utility scales and utility assessment . . . 591
16.4 Multiattribute Utility Functions . . . 593
Dominance . . . 594
Preference structure and multiattribute utility . . . 596
16.5 Decision Networks . . . 597

Representing a decision problem with a decision network . . . 598
Evaluating decision networks . . . 599
16.6 The Value of Information . . . 600
A simple example . . . 600
A general formula . . . 601
Properties of the value of information . . . 602
Implementing an information-gathering agent . . . 603
16.7 Decision-Theoretic Expert Systems . . . 604
16.8 Summary . . . 607
Bibliographical and Historical Notes . . . 607
Exercises . . . 609

17 Making Complex Decisions **613**
17.1 Sequential Decision Problems . . . 613
An example . . . 613
Optimality in sequential decision problems . . . 616
17.2 Value Iteration . . . 618
Utilities of states . . . 619
The value iteration algorithm . . . 620
Convergence of value iteration . . . 620
17.3 Policy Iteration . . . 624
17.4 Partially observable MDPs . . . 625
17.5 Decision-Theoretic Agents . . . 629
17.6 Decisions with Multiple Agents: Game Theory . . . 631
17.7 Mechanism Design . . . 640
17.8 Summary . . . 643
Bibliographical and Historical Notes . . . 644
Exercises . . . 646

VI Learning

18 Learning from Observations **649**
18.1 Forms of Learning . . . 649
18.2 Inductive Learning . . . 651
18.3 Learning Decision Trees . . . 653
Decision trees as performance elements . . . 653
Expressiveness of decision trees . . . 655
Inducing decision trees from examples . . . 655
Choosing attribute tests . . . 659
Assessing the performance of the learning algorithm . . . 660
Noise and overfitting . . . 661
Broadening the applicability of decision trees . . . 663
18.4 Ensemble Learning . . . 664
18.5 Why Learning Works: Computational Learning Theory . . . 668
How many examples are needed? . . . 669
Learning decision lists . . . 670
Discussion . . . 672
18.6 Summary . . . 673
Bibliographical and Historical Notes . . . 674

Exercises . . . 676

19 Knowledge in Learning **678**
19.1 A Logical Formulation of Learning . . . 678
Examples and hypotheses . . . 678
Current-best-hypothesis search . . . 680
Least-commitment search . . . 683
19.2 Knowledge in Learning . . . 686
Some simple examples . . . 687
Some general schemes . . . 688
19.3 Explanation-Based Learning . . . 690
Extracting general rules from examples . . . 691
Improving efficiency . . . 693
19.4 Learning Using Relevance Information . . . 694
Determining the hypothesis space . . . 695
Learning and using relevance information . . . 695
19.5 Inductive Logic Programming . . . 697
An example . . . 699
Top-down inductive learning methods . . . 701
Inductive learning with inverse deduction . . . 703
Making discoveries with inductive logic programming . . . 705
19.6 Summary . . . 707
Bibliographical and Historical Notes . . . 708
Exercises . . . 710

20 Statistical Learning Methods **712**
20.1 Statistical Learning . . . 712
20.2 Learning with Complete Data . . . 716
Maximum-likelihood parameter learning: Discrete models . . . 716
Naive Bayes models . . . 718
Maximum-likelihood parameter learning: Continuous models . . . 719
Bayesian parameter learning . . . 720
Learning Bayes net structures . . . 722
20.3 Learning with Hidden Variables: The EM Algorithm . . . 724
Unsupervised clustering: Learning mixtures of Gaussians . . . 725
Learning Bayesian networks with hidden variables . . . 727
Learning hidden Markov models . . . 731
The general form of the EM algorithm . . . 731
Learning Bayes net structures with hidden variables . . . 732
20.4 Instance-Based Learning . . . 733
Nearest-neighbor models . . . 733
Kernel models . . . 735
20.5 Neural Networks . . . 736
Units in neural networks . . . 737
Network structures . . . 738
Single layer feed-forward neural networks (perceptrons) . . . 740
Multilayer feed-forward neural networks . . . 744
Learning neural network structures . . . 748
20.6 Kernel Machines . . . 749

20.7 Case Study: Handwritten Digit Recognition . . . 752
20.8 Summary . . . 754
Bibliographical and Historical Notes . . . 755
Exercises . . . 759

21 Reinforcement Learning **763**
21.1 Introduction . . . 763
21.2 Passive Reinforcement Learning . . . 765
Direct utility estimation . . . 766
Adaptive dynamic programming . . . 767
Temporal difference learning . . . 767
21.3 Active Reinforcement Learning . . . 771
Exploration . . . 771
Learning an Action-Value Function . . . 775
21.4 Generalization in Reinforcement Learning . . . 777
Applications to game-playing . . . 780
Application to robot control . . . 780
21.5 Policy Search . . . 781
21.6 Summary . . . 784
Bibliographical and Historical Notes . . . 785
Exercises . . . 788

VII Communicating, perceiving, and acting

22 Communication **790**
22.1 Communication as Action . . . 790
Fundamentals of language . . . 791
The component steps of communication . . . 792
22.2 A Formal Grammar for a Fragment of English . . . 795
The Lexicon of $\mathcal{E}_0$ . . . 795
The Grammar of $\mathcal{E}_0$ . . . 796
22.3 Syntactic Analysis (Parsing) . . . 798
Efficient parsing . . . 800
22.4 Augmented Grammars . . . 806
Verb subcategorization . . . 808
Generative capacity of augmented grammars . . . 809
22.5 Semantic Interpretation . . . 810
The semantics of an English fragment . . . 811
Time and tense . . . 812
Quantification . . . 813
Pragmatic Interpretation . . . 815
Language generation with DCGs . . . 817
22.6 Ambiguity and Disambiguation . . . 818
Disambiguation . . . 820
22.7 Discourse Understanding . . . 821
Reference resolution . . . 821
The structure of coherent discourse . . . 823
22.8 Grammar Induction . . . 824
22.9 Summary . . . 826

Bibliographical and Historical Notes . . . 827
Exercises . . . 831

23 Probabilistic Language Processing **834**
23.1 Probabilistic Language Models . . . 834
Probabilistic context-free grammars . . . 836
Learning probabilities for PCFGs . . . 839
Learning rule structure for PCFGs . . . 840
23.2 Information Retrieval . . . 840
Evaluating IR systems . . . 842
IR refinements . . . 844
Presentation of result sets . . . 845
Implementing IR systems . . . 846
23.3 Information Extraction . . . 848
23.4 Machine Translation . . . 850
Machine translation systems . . . 852
Statistical machine translation . . . 853
Learning probabilities for machine translation . . . 856
23.5 Summary . . . 857
Bibliographical and Historical Notes . . . 858
Exercises . . . 861

24 Perception **863**
24.1 Introduction . . . 863
24.2 Image Formation . . . 865
Images without lenses: the pinhole camera . . . 865
Lens systems . . . 866
Light: the photometry of image formation . . . 867
Color: the spectrophotometry of image formation . . . 868
24.3 Early Image Processing Operations . . . 869
Edge detection . . . 870
Image segmentation . . . 872
24.4 Extracting Three-Dimensional Information . . . 873
Motion . . . 875
Binocular stereopsis . . . 876
Texture gradients . . . 879
Shading . . . 880
Contour . . . 881
24.5 Object Recognition . . . 885
Brightness-based recognition . . . 887
Feature-based recognition . . . 888
Pose Estimation . . . 890
24.6 Using Vision for Manipulation and Navigation . . . 892
24.7 Summary . . . 894
Bibliographical and Historical Notes . . . 895
Exercises . . . 898

25 Robotics **901**
25.1 Introduction . . . 901

25.2 Robot Hardware . . . 903
Sensors . . . 903
Effectors . . . 904
25.3 Robotic Perception . . . 907
Localization . . . 908
Mapping . . . 913
Other types of perception . . . 915
25.4 Planning to Move . . . 916
Configuration space . . . 916
Cell decomposition methods . . . 919
Skeletonization methods . . . 922
25.5 Planning uncertain movements . . . 923
Robust methods . . . 924
25.6 Moving . . . 926
Dynamics and control . . . 927
Potential field control . . . 929
Reactive control . . . 930
25.7 Robotic Software Architectures . . . 932
Subsumption architecture . . . 932
Three-layer architecture . . . 933
Robotic programming languages . . . 934
25.8 Application Domains . . . 935
25.9 Summary . . . 938
Bibliographical and Historical Notes . . . 939
Exercises . . . 942

VIII Conclusions

26 Philosophical Foundations **947**
26.1 Weak AI: Can Machines Act Intelligently? . . . 947
The argument from disability . . . 948
The mathematical objection . . . 949
The argument from informality . . . 950
26.2 Strong AI: Can Machines Really Think? . . . 952
The mind–body problem . . . 954
The "brain in a vat" experiment . . . 955
The brain prosthesis experiment . . . 956
The Chinese room . . . 958
26.3 The Ethics and Risks of Developing Artificial Intelligence . . . 960
26.4 Summary . . . 964
Bibliographical and Historical Notes . . . 964
Exercises . . . 967

27 AI: Present and Future **968**
27.1 Agent Components . . . 968
27.2 Agent Architectures . . . 970
27.3 Are We Going in the Right Direction? . . . 972

27.4 What if AI Does Succeed? 974

A Mathematical background **977**
A.1 Complexity Analysis and O() Notation 977
Asymptotic analysis 977
NP and inherently hard problems 978
A.2 Vectors, Matrices, and Linear Algebra 979
A.3 Probability Distributions 981
Bibliographical and Historical Notes 983

B Notes on Languages and Algorithms **984**
B.1 Defining Languages with Backus–Naur Form (BNF) 984
B.2 Describing Algorithms with Pseudocode 985
B.3 Online Help 985

Bibliography **987**

Index **1045**