

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



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