

Contents

I 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	17
1.4	The State of the Art	27
1.5	Risks and Benefits of AI	31
	Summary	34
	Bibliographical and Historical Notes	35
2	Intelligent Agents	36
2.1	Agents and Environments	36
2.2	Good Behavior: The Concept of Rationality	39
2.3	The Nature of Environments	42
2.4	The Structure of Agents	47
	Summary	60
	Bibliographical and Historical Notes	60
 II Problem-solving		
3	Solving Problems by Searching	63
3.1	Problem-Solving Agents	63
3.2	Example Problems	66
3.3	Search Algorithms	71
3.4	Uninformed Search Strategies	76
3.5	Informed (Heuristic) Search Strategies	84
3.6	Heuristic Functions	97
	Summary	104
	Bibliographical and Historical Notes	106
4	Search in Complex Environments	110
4.1	Local Search and Optimization Problems	110
4.2	Local Search in Continuous Spaces	119
4.3	Search with Nondeterministic Actions	122
4.4	Search in Partially Observable Environments	126
4.5	Online Search Agents and Unknown Environments	134
	Summary	141
	Bibliographical and Historical Notes	142
5	Adversarial Search and Games	146
5.1	Game Theory	146
5.2	Optimal Decisions in Games	148

5.3	Heuristic Alpha–Beta Tree Search	156
5.4	Monte Carlo Tree Search	161
5.5	Stochastic Games	164
5.6	Partially Observable Games	168
5.7	Limitations of Game Search Algorithms	173
	Summary	174
	Bibliographical and Historical Notes	175
6	Constraint Satisfaction Problems	180
6.1	Defining Constraint Satisfaction Problems	180
6.2	Constraint Propagation: Inference in CSPs	185
6.3	Backtracking Search for CSPs	191
6.4	Local Search for CSPs	197
6.5	The Structure of Problems	199
	Summary	203
	Bibliographical and Historical Notes	204
III	Knowledge, reasoning, and planning	
7	Logical Agents	208
7.1	Knowledge-Based Agents	209
7.2	The Wumpus World	210
7.3	Logic	214
7.4	Propositional Logic: A Very Simple Logic	217
7.5	Propositional Theorem Proving	222
7.6	Effective Propositional Model Checking	232
7.7	Agents Based on Propositional Logic	237
	Summary	246
	Bibliographical and Historical Notes	247
8	First-Order Logic	251
8.1	Representation Revisited	251
8.2	Syntax and Semantics of First-Order Logic	256
8.3	Using First-Order Logic	265
8.4	Knowledge Engineering in First-Order Logic	271
	Summary	277
	Bibliographical and Historical Notes	278
9	Inference in First-Order Logic	280
9.1	Propositional vs. First-Order Inference	280
9.2	Unification and First-Order Inference	282
9.3	Forward Chaining	286
9.4	Backward Chaining	293
9.5	Resolution	298
	Summary	309
	Bibliographical and Historical Notes	310

10 Knowledge Representation	314
10.1 Ontological Engineering	314
10.2 Categories and Objects	317
10.3 Events	322
10.4 Mental Objects and Modal Logic	326
10.5 Reasoning Systems for Categories	329
10.6 Reasoning with Default Information	333
Summary	337
Bibliographical and Historical Notes	338
11 Automated Planning	344
11.1 Definition of Classical Planning	344
11.2 Algorithms for Classical Planning	348
11.3 Heuristics for Planning	353
11.4 Hierarchical Planning	356
11.5 Planning and Acting in Nondeterministic Domains	365
11.6 Time, Schedules, and Resources	374
11.7 Analysis of Planning Approaches	378
Summary	379
Bibliographical and Historical Notes	380
IV Uncertain knowledge and reasoning	
12 Quantifying Uncertainty	385
12.1 Acting under Uncertainty	385
12.2 Basic Probability Notation	388
12.3 Inference Using Full Joint Distributions	395
12.4 Independence	397
12.5 Bayes' Rule and Its Use	399
12.6 Naive Bayes Models	402
12.7 The Wumpus World Revisited	404
Summary	407
Bibliographical and Historical Notes	408
13 Probabilistic Reasoning	412
13.1 Representing Knowledge in an Uncertain Domain	412
13.2 The Semantics of Bayesian Networks	414
13.3 Exact Inference in Bayesian Networks	427
13.4 Approximate Inference for Bayesian Networks	435
13.5 Causal Networks	449
Summary	453
Bibliographical and Historical Notes	454
14 Probabilistic Reasoning over Time	461
14.1 Time and Uncertainty	461
14.2 Inference in Temporal Models	465

14.3	Hidden Markov Models	473
14.4	Kalman Filters	479
14.5	Dynamic Bayesian Networks	485
	Summary	496
	Bibliographical and Historical Notes	497
15	Probabilistic Programming	500
15.1	Relational Probability Models	501
15.2	Open-Universe Probability Models	507
15.3	Keeping Track of a Complex World	514
15.4	Programs as Probability Models	519
	Summary	523
	Bibliographical and Historical Notes	524
16	Making Simple Decisions	528
16.1	Combining Beliefs and Desires under Uncertainty	528
16.2	The Basis of Utility Theory	529
16.3	Utility Functions	532
16.4	Multiattribute Utility Functions	540
16.5	Decision Networks	544
16.6	The Value of Information	547
16.7	Unknown Preferences	553
	Summary	557
	Bibliographical and Historical Notes	557
17	Making Complex Decisions	562
17.1	Sequential Decision Problems	562
17.2	Algorithms for MDPs	572
17.3	Bandit Problems	581
17.4	Partially Observable MDPs	588
17.5	Algorithms for Solving POMDPs	590
	Summary	595
	Bibliographical and Historical Notes	596
18	Multiagent Decision Making	599
18.1	Properties of Multiagent Environments	599
18.2	Non-Cooperative Game Theory	605
18.3	Cooperative Game Theory	626
18.4	Making Collective Decisions	632
	Summary	645
	Bibliographical and Historical Notes	646
V	Machine Learning	
19	Learning from Examples	651
19.1	Forms of Learning	651

19.2	Supervised Learning	653
19.3	Learning Decision Trees	657
19.4	Model Selection and Optimization	665
19.5	The Theory of Learning	672
19.6	Linear Regression and Classification	676
19.7	Nonparametric Models	686
19.8	Ensemble Learning	696
19.9	Developing Machine Learning Systems	704
	Summary	714
	Bibliographical and Historical Notes	715
20	Learning Probabilistic Models	721
20.1	Statistical Learning	721
20.2	Learning with Complete Data	724
20.3	Learning with Hidden Variables: The EM Algorithm	737
	Summary	746
	Bibliographical and Historical Notes	747
21	Deep Learning	750
21.1	Simple Feedforward Networks	751
21.2	Computation Graphs for Deep Learning	756
21.3	Convolutional Networks	760
21.4	Learning Algorithms	765
21.5	Generalization	768
21.6	Recurrent Neural Networks	772
21.7	Unsupervised Learning and Transfer Learning	775
21.8	Applications	782
	Summary	784
	Bibliographical and Historical Notes	785
22	Reinforcement Learning	789
22.1	Learning from Rewards	789
22.2	Passive Reinforcement Learning	791
22.3	Active Reinforcement Learning	797
22.4	Generalization in Reinforcement Learning	803
22.5	Policy Search	810
22.6	Apprenticeship and Inverse Reinforcement Learning	812
22.7	Applications of Reinforcement Learning	815
	Summary	818
	Bibliographical and Historical Notes	819
VI	Communicating, perceiving, and acting	
23	Natural Language Processing	823
23.1	Language Models	823
23.2	Grammar	833

23.3	Parsing	835
23.4	Augmented Grammars	841
23.5	Complications of Real Natural Language	845
23.6	Natural Language Tasks	849
	Summary	850
	Bibliographical and Historical Notes	851
24	Deep Learning for Natural Language Processing	856
24.1	Word Embeddings	856
24.2	Recurrent Neural Networks for NLP	860
24.3	Sequence-to-Sequence Models	864
24.4	The Transformer Architecture	868
24.5	Pretraining and Transfer Learning	871
24.6	State of the art	875
	Summary	878
	Bibliographical and Historical Notes	878
25	Computer Vision	881
25.1	Introduction	881
25.2	Image Formation	882
25.3	Simple Image Features	888
25.4	Classifying Images	895
25.5	Detecting Objects	899
25.6	The 3D World	901
25.7	Using Computer Vision	906
	Summary	919
	Bibliographical and Historical Notes	920
26	Robotics	925
26.1	Robots	925
26.2	Robot Hardware	926
26.3	What kind of problem is robotics solving?	930
26.4	Robotic Perception	931
26.5	Planning and Control	938
26.6	Planning Uncertain Movements	956
26.7	Reinforcement Learning in Robotics	958
26.8	Humans and Robots	961
26.9	Alternative Robotic Frameworks	968
26.10	Application Domains	971
	Summary	974
	Bibliographical and Historical Notes	975
VII	Conclusions	
27	Philosophy, Ethics, and Safety of AI	981
27.1	The Limits of AI	981

27.2	Can Machines Really Think?	984
27.3	The Ethics of AI	986
	Summary	1005
	Bibliographical and Historical Notes	1006
28	The Future of AI	1012
28.1	AI Components	1012
28.2	AI Architectures	1018
A	Mathematical Background	1023
A.1	Complexity Analysis and $O()$ Notation	1023
A.2	Vectors, Matrices, and Linear Algebra	1025
A.3	Probability Distributions	1027
	Bibliographical and Historical Notes	1029
B	Notes on Languages and Algorithms	1030
B.1	Defining Languages with Backus–Naur Form (BNF)	1030
B.2	Describing Algorithms with Pseudocode	1031
B.3	Online Supplemental Material	1032
	Bibliography	1033
	Index	1069