

Contents

Preface	xxi
I Agents in the World	1
1 Artificial Intelligence and Agents	3
1.1 What is Artificial Intelligence?	3
1.1.1 Artificial and Natural Intelligence	5
1.1.2 Natural Intelligence	8
1.2 A Brief History of Artificial Intelligence	9
1.2.1 Relationship to Other Disciplines	12
1.3 Agents Situated in Environments	14
1.4 Prototypical Applications	16
1.4.1 An Autonomous Delivery and Helping Robot	16
1.4.2 A Diagnostic Assistant	17
1.4.3 A Tutoring Agent	19
1.4.4 A Trading Agent	20
1.4.5 Smart Home	20
1.5 Agent Design Space	21
1.5.1 Modularity	22
1.5.2 Planning Horizon	23
1.5.3 Representation	24
1.5.4 Computational Limits	26
1.5.5 Learning	28
1.5.6 Uncertainty	29
1.5.7 Preference	31

1.5.8	Number of Agents	33
1.5.9	Interactivity	34
1.5.10	Interaction of the Dimensions	35
1.6	Designing Agents	37
1.6.1	Simplifying Environments and Simplifying Agents . . .	37
1.6.2	Tasks	39
1.6.3	Defining a Solution	40
1.6.4	Representations	42
1.7	Social Impact	44
1.8	Overview of the Book	45
1.9	Review	46
1.10	References and Further Reading	47
1.11	Exercises	48
2	Agent Architectures and Hierarchical Control	51
2.1	Agents and Environments	51
2.1.1	Controllers	53
2.1.2	Belief States	55
2.1.3	Agent Functions	56
2.2	Hierarchical Control	58
2.3	Designing Agents	65
2.3.1	Discrete, Continuous, and Hybrid	65
2.3.2	Choosing Agent Functions	67
2.3.3	Offline and Online Computation	68
2.4	Social Impact	71
2.5	Review	72
2.6	References and Further Reading	72
2.7	Exercises	73
II	Reasoning and Planning with Certainty	77
3	Searching for Solutions	79
3.1	Problem Solving as Search	79
3.2	State Spaces	81
3.3	Graph Searching	84
3.3.1	Formalizing Graph Searching	84
3.4	A Generic Searching Algorithm	86
3.5	Uninformed Search Strategies	88
3.5.1	Breadth-First Search	90
3.5.2	Depth-First Search	91
3.5.3	Iterative Deepening	96
3.5.4	Lowest-Cost-First Search	99
3.6	Informed (Heuristic) Search	101
3.6.1	A* Search	102

3.6.2	Branch and Bound	105
3.6.3	Designing a Heuristic Function	108
3.7	Pruning the Search Space	109
3.7.1	Cycle Pruning	109
3.7.2	Multiple-Path Pruning	110
3.7.3	Summary of Search Strategies	113
3.8	Search Refinements	114
3.8.1	Direction of Search	115
3.8.2	Dynamic Programming	117
3.9	Social Impact	120
3.10	Review	121
3.11	References and Further Reading	121
3.12	Exercises	122
4	Reasoning with Constraints	127
4.1	Variables and Constraints	127
4.1.1	Variables and Assignments	127
4.1.2	Constraints	131
4.1.3	Constraint Satisfaction Problems	132
4.2	Solving CSPs by Searching	133
4.3	Consistency Algorithms	136
4.4	Domain Splitting	141
4.5	Variable Elimination	143
4.6	Local Search	146
4.6.1	Iterative Best Improvement	148
4.6.2	Randomized Algorithms	149
4.6.3	Local Search Variants	150
4.6.4	Evaluating Randomized Algorithms	155
4.6.5	Random Restart	157
4.7	Population-Based Methods	158
4.8	Optimization	161
4.8.1	Systematic Methods for Discrete Optimization	164
4.8.2	Local Search for Optimization	167
4.8.3	Gradient Descent for Continuous Functions	167
4.9	Social Impact	170
4.10	Review	171
4.11	References and Further Reading	171
4.12	Exercises	172
5	Propositions and Inference	177
5.1	Propositions	177
5.1.1	Syntax of the Propositional Calculus	177
5.1.2	Semantics of the Propositional Calculus	178
5.2	Propositional Constraints	181
5.2.1	Clausal Form for CSPs	182

5.2.2	Exploiting Propositional Structure in Local Search	184
5.3	Propositional Definite Clauses	185
5.3.1	Queries and Answers	187
5.3.2	Proofs	188
5.4	Querying the User	195
5.5	Knowledge-Level Debugging	196
5.5.1	Incorrect Answers	197
5.5.2	Missing Answers	198
5.6	Proving by Contradiction	200
5.6.1	Horn Clauses	200
5.6.2	Assumables and Conflicts	201
5.6.3	Consistency-Based Diagnosis	202
5.6.4	Reasoning with Assumptions and Horn Clauses	204
5.7	Complete Knowledge Assumption	207
5.7.1	Non-Monotonic Reasoning	210
5.7.2	Proof Procedures for Negation as Failure	211
5.8	Abduction	214
5.9	Causal Models	217
5.10	Social Impact	219
5.11	Review	220
5.12	References and Further Reading	221
5.13	Exercises	222
6	Deterministic Planning	231
6.1	Representing States, Actions, and Goals	232
6.1.1	Explicit State-Space Representation	233
6.1.2	The STRIPS Representation	235
6.1.3	Feature-Based Representation of Actions	237
6.1.4	Initial States and Goals	238
6.2	Forward Planning	239
6.3	Regression Planning	241
6.4	Planning as a CSP	244
6.4.1	Action Features	247
6.5	Partial-Order Planning	248
6.6	Social Impact	252
6.7	Review	253
6.8	References and Further Reading	253
6.9	Exercises	254
III	Learning and Reasoning with Uncertainty	259
7	Supervised Machine Learning	261
7.1	Learning Issues	262
7.2	Supervised Learning Foundations	266

7.2.1	Evaluating Predictions	269
7.2.2	Point Estimates with No Input Features	276
7.2.3	Types of Errors	278
7.3	Basic Models for Supervised Learning	281
7.3.1	Learning Decision Trees	281
7.3.2	Linear Regression and Classification	288
7.4	Overfitting	297
7.4.1	Pseudocounts	301
7.4.2	Regularization	302
7.4.3	Cross Validation	304
7.5	Composite Models	307
7.5.1	Boosting	309
7.5.2	Gradient-Boosted Trees	311
7.6	Limitations	315
7.7	Social Impact	317
7.8	Review	319
7.9	References and Further Reading	319
7.10	Exercises	320
8	Neural Networks and Deep Learning	327
8.1	Feedforward Neural Networks	329
8.1.1	Parameter Learning	332
8.2	Improved Optimization	336
8.2.1	Momentum	339
8.2.2	RMS-Prop	339
8.2.3	Adam	340
8.2.4	Initialization	341
8.3	Improving Generalization	342
8.4	Convolutional Neural Networks	344
8.5	Neural Models for Sequences	350
8.5.1	Word Embeddings	350
8.5.2	Recurrent Neural Networks	354
8.5.3	Long Short-Term Memory	357
8.5.4	Attention and Transformers	360
8.5.5	Large Language Models	364
8.6	Other Neural Network Models	366
8.6.1	Autoencoders	366
8.6.2	Adversarial Networks	366
8.6.3	Diffusion Models	367
8.7	Social Impact	367
8.8	Review	369
8.9	References and Further Reading	370
8.10	Exercises	372

9	Reasoning with Uncertainty	375
9.1	Probability	375
9.1.1	Semantics of Probability	377
9.1.2	Conditional Probability	378
9.1.3	Expected Values	383
9.2	Independence	384
9.3	Belief Networks	385
9.3.1	Observations and Queries	387
9.3.2	Constructing Belief Networks	388
9.3.3	Representing Conditional Probabilities and Factors	394
9.4	Probabilistic Inference	404
9.5	Exact Probabilistic Inference	405
9.5.1	Recursive Conditioning	409
9.5.2	Variable Elimination for Belief Networks	413
9.5.3	Exploiting Structure and Compilation	416
9.6	Sequential Probability Models	418
9.6.1	Markov Chains	418
9.6.2	Hidden Markov Models	420
9.6.3	Algorithms for Monitoring and Smoothing	426
9.6.4	Dynamic Belief Networks	427
9.6.5	Time Granularity	428
9.6.6	Probabilistic Language Models	430
9.7	Stochastic Simulation	436
9.7.1	Sampling from a Single Variable	437
9.7.2	Forward Sampling	439
9.7.3	Rejection Sampling	440
9.7.4	Likelihood Weighting	441
9.7.5	Importance Sampling	443
9.7.6	Particle Filtering	445
9.7.7	Markov Chain Monte Carlo	447
9.8	Social Impact	449
9.9	Review	450
9.10	References and Further Reading	450
9.11	Exercises	451
10	Learning with Uncertainty	459
10.1	Probabilistic Learning	459
10.2	Bayesian Learning	460
10.2.1	Learning Probabilities	461
10.2.2	Probabilistic Classifiers	467
10.2.3	Probabilistic Learning of Decision Trees	471
10.2.4	Description Length	472
10.3	Unsupervised Learning	473
10.3.1	k -Means	473
10.3.2	Expectation Maximization for Soft Clustering	478

10.4	Learning Belief Networks	481
10.4.1	Hidden Variables	482
10.4.2	Missing Data	482
10.4.3	Structure Learning	484
10.4.4	General Case of Belief Network Learning	485
10.5	Social Impact	486
10.6	Review	486
10.7	References and Further Reading	487
10.8	Exercises	487
11	Causality	491
11.1	Probabilistic Causal Models	492
11.1.1	Do-notation	494
11.1.2	D-separation	495
11.2	Missing Data	497
11.3	Inferring Causality	500
11.3.1	Backdoor Criterion	501
11.3.2	Do-calculus	502
11.3.3	Front-Door Criterion	503
11.3.4	Simpson's Paradox	504
11.4	Instrumental Variables	506
11.5	Counterfactual Reasoning	508
11.6	Social Impact	510
11.7	Review	511
11.8	References and Further Reading	512
11.9	Exercises	512
IV	Planning and Acting with Uncertainty	515
12	Planning with Uncertainty	517
12.1	Preferences and Utility	518
12.1.1	Axioms for Rationality	518
12.1.2	Factored Utility	526
12.1.3	Prospect Theory	528
12.2	One-Off Decisions	530
12.2.1	Single-Stage Decision Networks	534
12.3	Sequential Decisions	536
12.3.1	Decision Networks	537
12.3.2	Policies	541
12.3.3	Optimizing Decision Networks using Search	543
12.3.4	Variable Elimination for Decision Networks	544
12.4	The Value of Information and Control	549
12.5	Decision Processes	552
12.5.1	Policies	557

16.3.1	Description Logic	718
16.3.2	Top-Level Ontologies	723
16.4	Social Impact	726
16.5	Review	727
16.6	References and Further Reading	727
16.7	Exercises	728
17	Relational Learning and Probabilistic Reasoning	731
17.1	From Relations to Features and Random Variables	732
17.2	Embedding-Based models	734
17.2.1	Learning a Binary Relation	734
17.2.2	Learning Knowledge Graphs	740
17.3	Learning Interdependence of Relations	743
17.3.1	Relational Probabilistic Models	743
17.3.2	Collective Classification and Crowd Sourcing	749
17.3.3	Language and Topic Models	750
17.3.4	Some Specific Representations	754
17.4	Existence and Identity Uncertainty	756
17.5	Social Impact	758
17.6	Review	759
17.7	References and Further Reading	759
17.8	Exercises	760
VI	The Big Picture	765
18	The Social Impact of Artificial Intelligence	767
18.1	The Digital Economy	768
18.2	Values and Bias	769
18.3	Human-Centred Artificial Intelligence	770
18.4	Work and Automation	773
18.5	Transportation	774
18.6	Sustainability	775
18.7	Ethics	778
18.8	Governance and Regulation	782
18.9	Review	783
18.10	Exercises	784
19	Retrospect and Prospect	785
19.1	Deploying AI	785
19.2	Agent Design Space Revisited	788
19.3	Looking Ahead	792
19.4	References and Further Reading	794
19.5	Exercises	795

Appendix A	Mathematical Preliminaries and Notation	797
A.1	Rolling Average	797
A.2	Discrete Mathematics	798
A.3	Functions, Factors, and Arrays	799
A.4	Relations and the Relational Algebra	800
Appendix B	Mapping to Open-Source Packages	803
B.1	Gradient-Boosted Trees	803
B.2	Deep Learning	804
	References	807
	Index of Algorithms	843
	Index	847

Artificial Intelligence: Foundations of Computational Agents is a book about the science of artificial intelligence (AI). AI is the study of intelligent computational agents. Our book is structured as a textbook but it is designed to be accessible to a wide audience.

We wrote this book because we are excited about the emergence of AI as an integrated science. As with any science being developed, AI has a coherent formal theory and a rambunctious experimental wing. Here we balance theory and experiment and show how to link them together intimately. We develop the science of AI together with its engineering applications. We believe the adage, “There is nothing so practical as a good theory.” The spirit of our approach is captured by the dictum, “Everything should be made as simple as possible, but not simpler.” We must build the science on solid foundations; we present the foundations, but only sketch, and give some examples of, the complexity required to build useful intelligent systems. Although the resulting systems will be complex, the foundations and the building blocks should be simple.

New to This Edition

This third edition results from extensive revision throughout the text. We have added three new chapters:

- Neural Networks and Deep Learning
- Causality
- The Social Impact of Artificial Intelligence

There is also a social impact section for each chapter, covering either beneficial applications or harmful impacts of AI, and often both. With the rise in the