

计 算 机 科 学 丛 书

# Artificial Intelligence A New Synthesis

## 人工智能

(英文版)

(美) Nils J. Nilsson 著



机械工业出版社  
China Machine Press



MORGAN  
KAUFMANN



计算机科学丛书

# 人工智能

(英文版)

## Artificial Intelligence A New Synthesis

(美) Nils J. Nilsson 著



机械工业出版社  
China Machine Press

Nils J. Nilsson: Artificial Intelligence, A New Synthesis.

Copyright © 1998 by Morgan Kaufmann Publishers, Inc.

Harcourt Asia Pte Ltd under special arrangement with Morgan Kaufmann authorizes China Machine Press to print and exclusively distribute this edition, which is the only authorized complete and unabridged reproduction of the latest American Edition published and priced for sale in China only, not including Hong Kong SAR and Taiwan.

Unauthorized export of this edition is a violation of the Copyright Act. Violation of this Law is subjected to Civil and Criminal penalties.

Harcourt亚洲公司在与Morgan Kaufmann公司的特殊合同下，授权机械工业出版社影印并独家发行本版本，该书为美国最新版未经删节的完整版本的复制品，专为在中国境内（不包含台湾地区和香港特别行政区）销售而定价并出版。

本版本未经授权的出口行为将违反版权法，违法者需承担一切民事和刑事责任。

版权所有，侵权必究。

本书版权登记号：图字：01-1999-2024

图书在版编目（CIP）数据

NJS/HK/03

人工智能：英文/(美)尼尔松(Nilsson, N. J.)著.-北京：机械工业出版社，1999.9  
(计算机科学丛书)

ISBN 7-111-07438-6

I.人… II.尼… III.人工智能-英文 IV.TP18

中国版本图书馆CIP数据核字(1999)第33788号

机械工业出版社（北京市西城区百万庄大街22号 邮政编码 100037）

北京牛山世兴印刷厂印刷·新华书店北京发行所发行

1999年9月第1版·2001年5月第2次印刷

787mm×1092mm1/16·33.5印张

印数：2 001-3 000册

定价：45.00元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

***For Scott and Ryan***

## **About the author**

**Nils J. Nilsson's** long and rich research career has contributed much to AI. His previous books, considered classics in the field, include **Learning Machines**, **Problem-Solving Methods in Artificial Intelligence**, **Logical Foundations of Artificial Intelligence**, and **Principles of Artificial Intelligence**. Dr. Nilsson is Kumagai Professor of Engineering, Emeritus, at Stanford University. He has served on the editorial boards of **Artificial Intelligence and Machine Learning** and as an area editor for the **Journal of the Association for Computing Machinery**. Former chairman of the Department of Computer Science at Stanford and Former Director of the SRI Artificial Intelligence Center, he is also a past president and fellow of the American Association for Artificial Intelligence.

# Preface

This introductory textbook employs a novel perspective from which to view topics in artificial intelligence (AI). I will consider a progression of AI systems or “agents,” each slightly more complex than its predecessor. I begin with elementary agents that react to sensed properties of their environments. Even such simple machines allow me to treat topics in machine vision, machine learning, and machine evolution. Then, by stages, I introduce techniques that allow agents to exploit information about the task environment that cannot be immediately sensed. Such knowledge can take the form of descriptive information about the state of the environment, iconic models of the environment, state-space graphs, and logical representations. Because the progression follows what plausibly might have been milestones in the evolution of animals, I have called this approach *evolutionary artificial intelligence*. I intend the book to be as much a proposal about how to think about AI as it is a description of AI techniques. Examples will be used to provide motivation and grounding.

Although I use agents to motivate and illustrate AI techniques, the techniques themselves have much broader application. Many ideas invented by AI researchers have been assimilated into computer science generally for applications in expert systems, natural language processing, human-machine interaction, information retrieval, graphics and image processing, data mining, and robotics (to name some examples). The agents theme serves to unify what might otherwise seem to be a collection of disparate topics.

Regarding coverage, my intention is to treat the middle ground between theory and applications. This middle ground is rich in important AI *ideas*, and in this book I try to motivate and explain the ideas that I think have lasting value in AI. (Being subject to the usual human frailties, I admit to possible errors of omission

and commission in selecting topics for inclusion.) Also, some subjects are treated in more depth than others—both because I thought some subjects more important and because I wanted to provide at least some examples of greater depth of exposition. Although some pseudocode algorithms are presented, the book is not an AI programming and implementation book. (Some “AI techniques” books are [Shoham 1994, Norvig 1992, Tracy & Bouthoorn 1997].) I do not give proofs of all of the important theoretical results, but I try to give intuitive arguments and citations to formal proofs. My goals are to present a modest-sized textbook for a one-semester introductory college course, to give the student and reader sufficient motivation and preparation to go on to more advanced AI courses, and to make the extensive literature on AI accessible.

A somewhat unconventional feature of the book is that machine learning is not treated as a separate topic; instead, various aspects of learning arise throughout the book. Neural nets and fundamental ideas about supervised learning are presented early; techniques for learning search heuristics and action policies are discussed in the chapters on search; rule learning, inductive logic programming, and explanation-based learning are treated toward the end of the chapters on logic; and learning plans is presented after discussing logic-based planning.

In my previous books, I included a “bibliographic and historical remarks” section at the end of each chapter. (Some readers may find those sections of some interest still.) I have not done so in the present book, both because AI history has now accumulated to such a great extent and because the longer text by [Russell & Norvig 1995] has already done such a thorough job in that regard. Instead, I include remarks and citations as appropriate throughout the text and provide some additional ones in discussion sections at the end of most chapters. The serious student who intends to specialize in AI research will want to consult many of the references. I hope the casual reader is not bothered by the many citations.

Sample exercises are included at the end of each chapter. They vary in difficulty from routine application of ideas presented in the book to mildly challenging. I expect that instructors will want to augment these problems with favorite ones of their own, including computer exercises and projects. (In keeping with my decision to concentrate on ideas instead of programs, I have not included any computer exercises or projects. Several good programming and project ideas can be found in texts devoted to AI programming techniques.)

The following typographical conventions are used in this book. Sans serif font is used for the names of actions and for “proto-English” sentences communicated among agents. SANS SERIF capitals are used for the names of computer languages, algorithms, and AI systems. Boldface capital letters, such as **W** and **X**, are used for vectors, matrices, and modal operators. Typewriter font is used for genetic programs, for expressions and subexpressions in the predicate calculus, and for

STRIPS rules and operators. Lowercase Greek letters are used for metavariables ranging over predicate-calculus expressions, subexpressions, and occasionally for substitutions. Uppercase Greek letters are used to denote sets of predicate-calculus formulas. Lowercase  $p$ 's are used to denote probabilities.

Students and researchers will find much helpful material about AI on the World Wide Web. I do not provide URLs here; any list written today would be incomplete and inaccurate within months. Use of one of the web search engines will quickly steer the reader to sites with sample applications, frequently asked questions, extensive bibliographies, research papers, programs, interactive demonstrations, announcements of workshops and conferences, homepages of researchers, and much more.

Material specifically in support of this book is provided on a Web page on the publisher's Web site at [www.mkp.com/nils](http://www.mkp.com/nils). If you discover any errors, please email them to the publisher at [aibugs@mkp.com](mailto:aibugs@mkp.com). Errata and clarifications can be found at <http://www.mkp.com/nils/clarified>.

My previous AI textbook, *Principles of Artificial Intelligence* (Morgan Kaufmann, 1980), is by now quite out of date, but some of the material in that book is still useful, and I have borrowed freely from it in preparing the present volume. Cross-checking against other AI textbooks (particularly [Russell & Norvig 1995, Rich & Knight 1991, Stefik 1995]) was also very helpful.

Students and teaching assistants in my Stanford courses on artificial intelligence and machine learning have already made several useful suggestions. I hope the following list includes most of them: Eyal Amir, David Andre, Scott Benson, George John, Steve Ketchpel, Ron Kohavi, Andrew Kosoresow, Ofer Matan, Karl Pflieger, and Charles Richards. Colleagues and reviewers at Stanford and elsewhere helped me learn what they already knew. Thanks to Helder Coelho, Oscar Firschein, Carolyn Hayes, Giorgio Ingargiola, Leslie Kaelbling, Daphne Koller, John Koza, Richard Korf, Pat Langley, John McCarthy, Bart Selman, Yoav Shoham, Devika Subramanian, Gheorghe Tecuci, and Michael Wellman. Special thanks go to Cheri Palmer, my production editor at Morgan Kaufmann, who kept me on schedule, cheerfully accepted my endless changes, and worked extra hard to meet a difficult publication date. Work on this book was carried on in the Robotics Laboratory of Stanford's Department of Computer Science and at the Santa Fe Institute. Continuing research support by the National Science Foundation is gratefully acknowledged.



# Contents

	<b>Preface</b>	<b>XVII</b>
<b>1</b>	<b><i>Introduction</i></b>	<b>1</b>
1.1	<b>What Is AI?</b>	<b>1</b>
1.2	<b>Approaches to Artificial Intelligence</b>	<b>6</b>
1.3	<b>Brief History of AI</b>	<b>8</b>
1.4	<b>Plan of the Book</b>	<b>11</b>
1.5	<b>Additional Readings and Discussion</b>	<b>14</b>
	<b>Exercises</b>	<b>17</b>
<hr/>		
<b>I</b>	<b>Reactive Machines</b>	<b>19</b>
<b>2</b>	<b><i>Stimulus-Response Agents</i></b>	<b>21</b>
2.1	<b>Perception and Action</b>	<b>21</b>
2.1.1	<b>Perception</b>	<b>24</b>
2.1.2	<b>Action</b>	<b>24</b>
2.1.3	<b>Boolean Algebra</b>	<b>25</b>
2.1.4	<b>Classes and Forms of Boolean Functions</b>	<b>26</b>

<b>2.2 Representing and Implementing Action Functions</b>	<b>27</b>
2.2.1 Production Systems	27
2.2.2 Networks	29
2.2.3 The Subsumption Architecture	32
<b>2.3 Additional Readings and Discussion</b>	<b>33</b>
Exercises	34

## **Neural Networks** **37**

<b>3.1 Introduction</b>	<b>37</b>
<b>3.2 Training Single TLUs</b>	<b>38</b>
3.2.1 TLU Geometry	38
3.2.2 Augmented Vectors	39
3.2.3 Gradient Descent Methods	39
3.2.4 The Widrow-Hoff Procedure	41
3.2.5 The Generalized Delta Procedure	41
3.2.6 The Error-Correction Procedure	43
<b>3.3 Neural Networks</b>	<b>44</b>
3.3.1 Motivation	44
3.3.2 Notation	45
3.3.3 The Backpropagation Method	46
3.3.4 Computing Weight Changes in the Final Layer	48
3.3.5 Computing Changes to the Weights in Intermediate Layers	48
<b>3.4 Generalization, Accuracy, and Overfitting</b>	<b>51</b>
<b>3.5 Additional Readings and Discussion</b>	<b>54</b>
Exercises	55

## **Machine Evolution** **59**

<b>4.1 Evolutionary Computation</b>	<b>59</b>
<b>4.2 Genetic Programming</b>	<b>60</b>
4.2.1 Program Representation in GP	60

---

4.2.2	The GP Process	62
4.2.3	Evolving a Wall-Following Robot	65
<b>4.3</b>	<b>Additional Readings and Discussion</b>	<b>69</b>
	<b>Exercises</b>	<b>69</b>
<b>5</b>	<b>State Machines</b>	<b>71</b>
5.1	Representing the Environment by Feature Vectors	71
5.2	Elman Networks	73
5.3	Iconic Representations	74
5.4	Blackboard Systems	77
5.5	Additional Readings and Discussion	80
	<b>Exercises</b>	<b>80</b>
<b>6</b>	<b>Robot Vision</b>	<b>85</b>
6.1	Introduction	85
6.2	Steering an Automobile	86
6.3	Two Stages of Robot Vision	88
6.4	Image Processing	91
6.4.1	Averaging	91
6.4.2	Edge Enhancement	93
6.4.3	Combining Edge Enhancement with Averaging	96
6.4.4	Region Finding	97
6.4.5	Using Image Attributes Other Than Intensity	101
6.5	Scene Analysis	102
6.5.1	Interpreting Lines and Curves in the Image	103
6.5.2	Model-Based Vision	106
6.6	Stereo Vision and Depth Information	108
6.7	Additional Readings and Discussion	110
	<b>Exercises</b>	<b>111</b>

---

## **II Search in State Spaces** **115**

### **7 Agents That Plan** **117**

- 7.1 Memory Versus Computation** **117**
- 7.2 State-Space Graphs** **118**
- 7.3 Searching Explicit State Spaces** **121**
- 7.4 Feature-Based State Spaces** **122**
- 7.5 Graph Notation** **124**
- 7.6 Additional Readings and Discussion** **125**
- Exercises** **126**

### **8 Uninformed Search** **129**

- 8.1 Formulating the State Space** **129**
- 8.2 Components of Implicit State-Space Graphs** **130**
- 8.3 Breadth-First Search** **131**
- 8.4 Depth-First or Backtracking Search** **133**
- 8.5 Iterative Deepening** **135**
- 8.6 Additional Readings and Discussion** **136**
- Exercises** **137**

### **9 Heuristic Search** **139**

- 9.1 Using Evaluation Functions** **139**
- 9.2 A General Graph-Searching Algorithm** **141**
  - 9.2.1 Algorithm A\*** **142**
  - 9.2.2 Admissibility of A\*** **145**
  - 9.2.3 The Consistency (or Monotone) Condition** **150**
  - 9.2.4 Iterative-Deepening A\*** **153**
  - 9.2.5 Recursive Best-First Search** **154**

---

9.3	Heuristic Functions and Search Efficiency	155
9.4	Additional Readings and Discussion	160
	Exercises	160
<b>10</b>	<b><i>Planning, Acting, and Learning</i></b>	<b>163</b>
10.1	The Sense/Plan/Act Cycle	163
10.2	Approximate Search	165
10.2.1	Island-Driven Search	166
10.2.2	Hierarchical Search	167
10.2.3	Limited-Horizon Search	169
10.2.4	Cycles	170
10.2.5	Building Reactive Procedures	170
10.3	Learning Heuristic Functions	172
10.3.1	Explicit Graphs	172
10.3.2	Implicit Graphs	173
10.4	Rewards Instead of Goals	175
10.5	Additional Readings and Discussion	177
	Exercises	178
<b>11</b>	<b><i>Alternative Search Formulations and Applications</i></b>	<b>181</b>
11.1	Assignment Problems	181
11.2	Constructive Methods	183
11.3	Heuristic Repair	187
11.4	Function Optimization	189
	Exercises	192
<b>12</b>	<b><i>Adversarial Search</i></b>	<b>195</b>
12.1	Two-Agent Games	195
12.2	The Minimax Procedure	197

---

<b>12.3</b>	<b>The Alpha-Beta Procedure</b>	202
<b>12.4</b>	<b>The Search Efficiency of the Alpha-Beta Procedure</b>	207
<b>12.5</b>	<b>Other Important Matters</b>	208
<b>12.6</b>	<b>Games of Chance</b>	208
<b>12.7</b>	<b>Learning Evaluation Functions</b>	210
<b>12.8</b>	<b>Additional Readings and Discussion</b>	212
	<b>Exercises</b>	213

---

### **III Knowledge Representation and Reasoning** 215

<b>13</b>	<b><i>The Propositional Calculus</i></b>	<b>217</b>
<b>13.1</b>	<b>Using Constraints on Feature Values</b>	217
<b>13.2</b>	<b>The Language</b>	219
<b>13.3</b>	<b>Rules of Inference</b>	220
<b>13.4</b>	<b>Definition of Proof</b>	221
<b>13.5</b>	<b>Semantics</b>	222
	13.5.1 Interpretations	222
	13.5.2 The Propositional Truth Table	223
	13.5.3 Satisfiability and Models	224
	13.5.4 Validity	224
	13.5.5 Equivalence	225
	13.5.6 Entailment	225
<b>13.6</b>	<b>Soundness and Completeness</b>	226
<b>13.7</b>	<b>The PSAT Problem</b>	227
<b>13.8</b>	<b>Other Important Topics</b>	228
	13.8.1 Language Distinctions	228
	13.8.2 Metatheorems	228
	13.8.3 Associative Laws	229



---

13.8.4	Distributive Laws	229
	<b>Exercises</b>	229
<b>14</b>	<b><i>Resolution in the Propositional Calculus</i></b>	<b>231</b>
14.1	<b>A New Rule of Inference: Resolution</b>	231
14.1.1	Clauses as wffs	231
14.1.2	Resolution on Clauses	231
14.1.3	Soundness of Resolution	232
14.2	<b>Converting Arbitrary wffs to Conjunctions of Clauses</b>	232
14.3	<b>Resolution Refutations</b>	233
14.4	<b>Resolution Refutation Search Strategies</b>	235
14.4.1	Ordering Strategies	235
14.4.2	Refinement Strategies	236
14.5	<b>Horn Clauses</b>	237
	<b>Exercises</b>	238
<b>15</b>	<b><i>The Predicate Calculus</i></b>	<b>239</b>
15.1	<b>Motivation</b>	239
15.2	<b>The Language and Its Syntax</b>	240
15.3	<b>Semantics</b>	241
15.3.1	Worlds	241
15.3.2	Interpretations	242
15.3.3	Models and Related Notions	243
15.3.4	Knowledge	244
15.4	<b>Quantification</b>	245
15.5	<b>Semantics of Quantifiers</b>	246
15.5.1	Universal Quantifiers	246
15.5.2	Existential Quantifiers	247
15.5.3	Useful Equivalences	247
15.5.4	Rules of Inference	247

<b>15.6</b>	<b>Predicate Calculus as a Language for Representing Knowledge</b>	<b>248</b>
15.6.1	Conceptualizations	248
15.6.2	Examples	248
<b>15.7</b>	<b>Additional Readings and Discussion</b>	<b>250</b>
	<b>Exercises</b>	<b>250</b>

## **16**     ***Resolution in the Predicate Calculus***     **253**

<b>16.1</b>	<b>Unification</b>	<b>253</b>
<b>16.2</b>	<b>Predicate-Calculus Resolution</b>	<b>256</b>
<b>16.3</b>	<b>Completeness and Soundness</b>	<b>257</b>
<b>16.4</b>	<b>Converting Arbitrary wffs to Clause Form</b>	<b>257</b>
<b>16.5</b>	<b>Using Resolution to Prove Theorems</b>	<b>260</b>
<b>16.6</b>	<b>Answer Extraction</b>	<b>261</b>
<b>16.7</b>	<b>The Equality Predicate</b>	<b>262</b>
<b>16.8</b>	<b>Additional Readings and Discussion</b>	<b>265</b>
	<b>Exercises</b>	<b>265</b>

## **17**     ***Knowledge-Based Systems***     **269**

<b>17.1</b>	<b>Confronting the Real World</b>	<b>269</b>
<b>17.2</b>	<b>Reasoning Using Horn Clauses</b>	<b>270</b>
<b>17.3</b>	<b>Maintenance in Dynamic Knowledge Bases</b>	<b>275</b>
<b>17.4</b>	<b>Rule-Based Expert Systems</b>	<b>280</b>
<b>17.5</b>	<b>Rule Learning</b>	<b>286</b>
17.5.1	Learning Propositional Calculus Rules	286
17.5.2	Learning First-Order Logic Rules	291
17.5.3	Explanation-Based Generalization	295
<b>17.6</b>	<b>Additional Readings and Discussion</b>	<b>297</b>
	<b>Exercises</b>	<b>298</b>

---

<b>18</b>	<b><i>Representing Commonsense Knowledge</i></b>	<b>301</b>
18.1	<b>The Commonsense World</b>	301
18.1.1	What Is Commonsense Knowledge?	301
18.1.2	Difficulties in Representing Commonsense Knowledge	303
18.1.3	The Importance of Commonsense Knowledge	304
18.1.4	Research Areas	305
18.2	<b>Time</b>	306
18.3	<b>Knowledge Representation by Networks</b>	308
18.3.1	Taxonomic Knowledge	308
18.3.2	Semantic Networks	309
18.3.3	Nonmonotonic Reasoning in Semantic Networks	309
18.3.4	Frames	312
18.4	<b>Additional Readings and Discussion</b>	313
	<b>Exercises</b>	314
<b>19</b>	<b><i>Reasoning with Uncertain Information</i></b>	<b>317</b>
19.1	<b>Review of Probability Theory</b>	317
19.1.1	Fundamental Ideas	317
19.1.2	Conditional Probabilities	320
19.2	<b>Probabilistic Inference</b>	323
19.2.1	A General Method	323
19.2.2	Conditional Independence	324
19.3	<b>Bayes Networks</b>	325
19.4	<b>Patterns of Inference in Bayes Networks</b>	328
19.5	<b>Uncertain Evidence</b>	329
19.6	<b>D-Separation</b>	330
19.7	<b>Probabilistic Inference in Polytrees</b>	332
19.7.1	Evidence Above	332
19.7.2	Evidence Below	334