

# Preface

*“Only a small community has concentrated on general intelligence. No one has tried to make a thinking machine . . .*

*The bottom line is that we really haven’t progressed too far toward a truly intelligent machine. We have collections of dumb specialists in small domains; the true majesty of general intelligence still awaits our attack. . . .*

*We have got to get back to the deepest questions of AI and general intelligence...”*

– Marvin Minsky

as interviewed in *Hal’s Legacy*, edited by David Stork, 2000.

Our goal in creating this edited volume has been to fill an apparent gap in the scientific literature, by providing a coherent presentation of a body of contemporary research that, in spite of its integral importance, has hitherto kept a very low profile within the scientific and intellectual community. This body of work has not been given a name before; in this book we christen it “Artificial General Intelligence” (AGI). What distinguishes AGI work from run-of-the-mill “artificial intelligence” research is that it is explicitly focused on engineering general intelligence in the short term. We have been active researchers in the AGI field for many years, and it has been a pleasure to gather together papers from our colleagues working on related ideas from their own perspectives. In the Introduction we give a conceptual overview of the AGI field, and also summarize and interrelate the key ideas of the papers in the subsequent chapters.

Of course, “general intelligence” does not mean exactly the same thing to all researchers. In fact it is not a fully well-defined term, and one of the issues raised in the papers contained here is how to define general intelligence in a way that provides maximally useful guidance to practical AI work. But,

nevertheless, there is a clear qualitative meaning to the term. What is meant by AGI is, loosely speaking, AI systems that possess a reasonable degree of self-understanding and autonomous self-control, and have the ability to solve a variety of complex problems in a variety of contexts, and to learn to solve new problems that they didn't know about at the time of their creation. A marked distinction exists between practical AGI work and, on the other hand:

- Pragmatic but specialized “narrow AI” research which is aimed at creating programs carrying out specific tasks like playing chess, diagnosing diseases, driving cars and so forth (most contemporary AI work falls into this category.)
- Purely theoretical AI research, which is aimed at clarifying issues regarding the nature of intelligence and cognition, but doesn't involve technical details regarding actually realizing artificially intelligent software.

Some of the papers presented here come close to the latter (purely theoretical) category, but we have selected them because the theoretical notions they contain seem likely to lead to such technical details in the medium-term future, and/or resonate very closely with the technical details of AGI designs proposed by other authors.

The audience we intend to reach includes the AI community, and also the broader community of scientists and students in related fields such as philosophy, neuroscience, linguistics, psychology, biology, sociology, anthropology and engineering. Significantly more so than narrow AI, AGI is interdisciplinary in nature, and a full appreciation of the general intelligence problem and its various potential solutions requires one to take a wide variety of different perspectives.

Not all significant AGI researchers are represented in these pages, but we have sought to bring together a multiplicity of perspectives, including many that disagree with our own. Bringing a diverse body of AGI research together in a single volume reveals the common themes among various researchers' work, and makes clear what the big open questions are in this vital and critical area of research. It is our hope that this book will interest more researchers and students in pursuing AGI research themselves, thus aiding in the progress of science.

In the three years that this book has been in the making, we have noticed a significant increase in interest in AGI-related research within the academic AI community, including a number of small conference workshops with titles related to “Human-Level Intelligence.” We consider this challenge to the overwhelming dominance of narrow-AI an extremely positive move; however, we submit that “Artificial General Intelligence” is a more sensible way to conceptualize the problem than “Human-Level Intelligence.” The AGI systems and approaches described in these pages are not necessarily oriented towards emulating the human brain; and given the heterogeneity of the human mind/brain and its highly various levels of competence at various sorts of tasks, it seems very difficult to define “Human-Level Intelligence” in any way that is generally

applicable to AI systems that are fundamentally non-human-like in conception. On the other hand, the work of Hutter and Schmidhuber reported here provides a reasonable, abstract mathematical characterization of general intelligence which, while not in itself providing a practical approach to AGI design and engineering, at least provides a conceptually meaningful formalization of the ultimate goal of AGI work.

The grand goal of AGI remains mostly unrealized, and how long it will be until this situation is remedied remains uncertain. Among scientists who believe in the fundamental possibility of strong AI, the most optimistic serious estimates we have heard are in the range of 5-10 years, and the most pessimistic are in the range of centuries. While none of the articles contained here purports to present a complete solution to the AGI problem, we believe that they collectively embody meaningful conceptual progress, and indicate clearly that the direct pursuit of AGI is an endeavor worthy of significant research attention.



# Contents

## Contemporary Approaches to Artificial General Intelligence

*Cassio Pennachin, Ben Goertzel*

- 1 A Brief History of AGI . . . . . 1
  - 1.1 Some Historical AGI-Related Projects . . . . . 2
- 2 What Is Intelligence? . . . . . 6
  - 2.1 The Psychology of Intelligence . . . . . 6
  - 2.2 The Turing Test . . . . . 8
  - 2.3 A Control Theory Approach to Defining Intelligence . . . . . 8
  - 2.4 Efficient Intelligence . . . . . 10
- 3 The Abstract Theory of General Intelligence . . . . . 11
- 4 Toward a Pragmatic Logic . . . . . 15
- 5 Emulating the Human Brain . . . . . 17
- 6 Emulating the Human Mind . . . . . 19
- 7 Creating Intelligence by Creating Life . . . . . 22
- 8 The Social Nature of Intelligence . . . . . 24
- 9 Integrative Approaches . . . . . 26
- 10 The Outlook for AGI . . . . . 27
- Acknowledgments . . . . . 28
- References . . . . . 28

## The Logic of Intelligence

*Pei Wang*

- 1 Intelligence and Logic . . . . . 31
  - 1.1 To Define Intelligence . . . . . 31
  - 1.2 A Working Definition of Intelligence . . . . . 33
  - 1.3 Comparison With Other Definitions . . . . . 35
  - 1.4 Logic and Reasoning Systems . . . . . 40
- 2 The Components of NARS . . . . . 43
  - 2.1 Experience-Grounded Semantics . . . . . 43
  - 2.2 Inheritance Statement . . . . . 45
  - 2.3 Categorical Language . . . . . 47
  - 2.4 Syllogistic Inference Rules . . . . . 48
  - 2.5 Controlled Concurrency in Dynamic Memory . . . . . 50
- 3 The Properties of NARS . . . . . 52
  - 3.1 Reasonable Solutions . . . . . 52

3.2	Unified Uncertainty Processing .....	53
3.3	NARS as a Parallel and Distributed Network .....	54
3.4	Resources Competition .....	56
3.5	Flexible Behaviors .....	57
3.6	Autonomy and Creativity .....	58
4	Conclusions .....	60
	References .....	60

**The Novamente Artificial Intelligence Engine**

*Ben Goertzel, Cassio Pennachin*

1	Introduction .....	63
1.1	The Novamente AGI System .....	64
1.2	Novamente for Knowledge Management and Data Analysis .....	65
2	Enabling Software Technologies .....	67
2.1	A Distributed Software Architecture for Integrative AI .....	68
2.2	Database Integration and Knowledge Integration .....	70
3	What Is Artificial General Intelligence? .....	72
3.1	What Is General Intelligence? .....	73
3.2	The Integrative Approach to AGI .....	75
3.3	Experiential Interactive Learning and Adaptive Self-modification .....	77
4	The Psynet Model of Mind .....	80
5	The Novamente AGI Design .....	83
5.1	An Integrative Knowledge Representation .....	84
5.2	The Mind OS .....	88
5.3	Atom Types .....	91
5.4	Novamente Maps .....	94
5.5	Mind Agents .....	95
5.6	Map Dynamics .....	96
5.7	Functional Specialization .....	99
5.8	Novamente and the Human Brain .....	100
5.9	Emergent Structures .....	102
6	Interacting with Humans and Data Stores .....	104
6.1	Data Sources .....	105
6.2	Knowledge Encoding .....	106
6.3	Querying .....	107
6.4	Formal Language Queries .....	108
6.5	Conversational Interaction .....	109
6.6	Report Generation .....	109
6.7	Active Collaborative Filtering and User Modeling .....	110
7	Example Novamente AI Processes .....	110
7.1	Probabilistic Inference .....	112
7.2	Nonlinear-Dynamical Attention Allocation .....	115
7.3	Importance Updating .....	116
7.4	Schema and Predicate Learning .....	117
7.5	Pattern Mining .....	120

7.6 Natural Language Processing . . . . . 122  
 8 Conclusion . . . . . 124  
 Appendix: Novamente Applied to Bioinformatic Pattern Mining . . . . . 125  
 References . . . . . 127

**Essentials of General Intelligence:  
 The Direct Path to Artificial General Intelligence**

*Peter Voss*

1 Introduction . . . . . 131  
 2 General Intelligence . . . . . 131  
   2.1 Core Requirements for General Intelligence . . . . . 133  
   2.2 Advantages of Intelligence Being General . . . . . 134  
 3 Shortcuts to AGI . . . . . 135  
 4 Foundational Cognitive Capabilities . . . . . 142  
 5 An AGI in the Making . . . . . 144  
   5.1 AGI Engine Architecture and Design Features . . . . . 145  
 6 From Algorithms to General Intelligence . . . . . 147  
   6.1 Sample Test Domains for Initial Performance Criteria . . . . . 148  
   6.2 Towards Increased Intelligence . . . . . 149  
 7 Other Research . . . . . 150  
 8 Fast-track AGI: Why So Rare? . . . . . 152  
 9 Conclusion . . . . . 155  
 References . . . . . 156

**Artificial Brains**

*Hugo de Garis*

1 Introduction . . . . . 159  
 2 Evolvable Hardware . . . . . 161  
   2.1 Neural Network Models . . . . . 162  
 3 The CAM-Brain Machine (CBM) . . . . . 166  
   3.1 Evolved Modules . . . . . 167  
   3.2 The Kitten Robot “Robokitty” . . . . . 168  
 4 Short- and Long-Term Future . . . . . 171  
 5 Postscript – July 2002 . . . . . 172  
 References . . . . . 174

**The New AI: General & Sound & Relevant for Physics**

*Jürgen Schmidhuber*

1 Introduction . . . . . 175  
 2 More Formally . . . . . 176  
 3 Prediction Using a Universal Algorithmic Prior Based on the  
   Shortest Way of Describing Objects . . . . . 177  
 4 Super Omegas and Generalizations of Kolmogorov Complexity &  
   Algorithmic Probability . . . . . 179  
 5 Computable Predictions Through the Speed Prior Based on the  
   Fastest Way of Describing Objects . . . . . 181

6	Speed Prior-Based Predictions for Our Universe . . . . .	182
7	Optimal Rational Decision Makers . . . . .	184
8	Optimal Universal Search Algorithms . . . . .	185
9	Optimal Ordered Problem Solver (OOPS) . . . . .	186
10	OOPS-Based Reinforcement Learning . . . . .	190
11	The Gödel Machine . . . . .	191
12	Conclusion . . . . .	192
13	Acknowledgments . . . . .	194
	References . . . . .	194

**Gödel Machines: Fully Self-Referential Optimal Universal Self-improvers**

*Jürgen Schmidhuber*

1	Introduction and Outline . . . . .	199
2	Basic Overview, Relation to Previous Work, and Limitations . . . . .	200
	2.1 Notation and Set-up . . . . .	201
	2.2 Basic Idea of Gödel Machine . . . . .	203
	2.3 Proof Techniques and an $O()$ -optimal Initial Proof Searcher. . . . .	203
	2.4 Relation to Hutter’s Previous Work . . . . .	204
	2.5 Limitations of Gödel Machines . . . . .	205
3	Essential Details of One Representative Gödel Machine . . . . .	206
	3.1 Proof Techniques . . . . .	206
4	Global Optimality Theorem . . . . .	212
	4.1 Alternative Relaxed Target Theorem . . . . .	212
5	Bias-Optimal Proof Search (BIOPS) . . . . .	213
	5.1 How a Surviving Proof Searcher May Use BIOPS to Solve Remaining Proof Search Tasks . . . . .	214
6	Discussion & Additional Relations to Previous Work . . . . .	215
	6.1 Possible Types of Gödel Machine Self-improvements . . . . .	215
	6.2 Example Applications . . . . .	217
	6.3 Probabilistic Gödel Machine Hardware . . . . .	217
	6.4 More Relations to Previous Work on Less General Self-improving Machines . . . . .	218
	6.5 Are Humans Probabilistic Gödel Machines? . . . . .	220
	6.6 Gödel Machines and Consciousness . . . . .	221
	6.7 Frequently Asked Questions . . . . .	221
7	Conclusion . . . . .	222
8	Acknowledgments . . . . .	223
	References . . . . .	223

**Universal Algorithmic Intelligence: A Mathematical Top→Down Approach**

*Marcus Hutter*

1	Introduction . . . . .	227
2	Agents in Known Probabilistic Environments . . . . .	230



2.1	The Cybernetic Agent Model . . . . .	230
2.2	Strings . . . . .	232
2.3	AI Model for Known Deterministic Environment . . . . .	232
2.4	AI Model for Known Prior Probability . . . . .	233
2.5	Probability Distributions . . . . .	235
2.6	Explicit Form of the $AI\mu$ Model . . . . .	236
2.7	Factorizable Environments . . . . .	238
2.8	Constants and Limits . . . . .	239
2.9	Sequential Decision Theory . . . . .	240
3	Universal Sequence Prediction . . . . .	241
3.1	Introduction . . . . .	241
3.2	Algorithmic Information Theory . . . . .	242
3.3	Uncertainty & Probabilities . . . . .	243
3.4	Algorithmic Probability & Universal Induction . . . . .	244
3.5	Loss Bounds & Pareto Optimality . . . . .	245
4	The Universal Algorithmic Agent AIXI . . . . .	246
4.1	The Universal $AI\xi$ Model . . . . .	246
4.2	On the Optimality of AIXI . . . . .	249
4.3	Value Bounds and Separability Concepts . . . . .	251
4.4	Pareto Optimality of $AI\xi$ . . . . .	254
4.5	The Choice of the Horizon . . . . .	255
4.6	Outlook . . . . .	257
4.7	Conclusions . . . . .	258
5	Important Problem Classes . . . . .	259
5.1	Sequence Prediction (SP) . . . . .	259
5.2	Strategic Games (SG) . . . . .	261
5.3	Function Minimization (FM) . . . . .	265
5.4	Supervised Learning from Examples (EX) . . . . .	269
5.5	Other Aspects of Intelligence . . . . .	271
6	Time-Bounded AIXI Model . . . . .	272
6.1	Time-Limited Probability Distributions . . . . .	273
6.2	The Idea of the Best Vote Algorithm . . . . .	275
6.3	Extended Chronological Programs . . . . .	276
6.4	Valid Approximations . . . . .	276
6.5	Effective Intelligence Order Relation . . . . .	277
6.6	The Universal Time-Bounded $AIXI^{tl}$ Agent . . . . .	277
6.7	Limitations and Open Questions . . . . .	278
6.8	Remarks . . . . .	279
7	Discussion . . . . .	280
7.1	General Remarks . . . . .	280
7.2	Outlook & Open Questions . . . . .	282
7.3	The Big Questions . . . . .	283
7.4	Conclusions . . . . .	284
	Annotated Bibliography . . . . .	285
	References . . . . .	287

**Program Search as a Path to Artificial General Intelligence**

*Lukasz Kaiser*

1 Intelligence and the Search for Programs . . . . . 291

2 Theoretical Results . . . . . 294

    2.1 Program Search in the Standard AI Model . . . . . 295

    2.2 Self-improving Program Search . . . . . 296

    2.3 Discussion of Efficiency Definitions . . . . . 298

3 Convenient Model of Computation . . . . . 299

    3.1 Extended Program Notation . . . . . 306

    3.2 Compiling Typed Rewriting Systems . . . . . 311

4 Reasoning Using Games . . . . . 314

    4.1 Reason and Search Game for Terms . . . . . 318

5 Conclusions . . . . . 324

References . . . . . 325

**The Natural Way to Artificial Intelligence**

*Vladimir G. Red'ko*

1 Introduction . . . . . 327

2 The Epistemological Problem . . . . . 328

3 Approaches to the Theory of Evolutionary Origin of Human  
Intelligence . . . . . 330

    3.1 “Intelligent Inventions” of Biological Evolution . . . . . 331

    3.2 Methodological Approaches . . . . . 334

    3.3 Role of Investigations of “Artificial Life” and “Simulation of  
Adaptive Behavior” . . . . . 337

4 Two Models . . . . . 338

    4.1 Alife Model of Evolutionary Emergence of Purposeful  
Adaptive Behavior . . . . . 338

    4.2 Model of Evolution of Web Agents . . . . . 343

5 Towards the Implementation of Higher Cognitive Abilities . . . . . 347

6 Conclusion . . . . . 349

7 Acknowledgements . . . . . 349

References . . . . . 349

**3D Simulation: the Key to A.I.**

*Keith A. Hoyes*

1 Introduction . . . . . 353

2 Pillars of Intelligence . . . . . 354

    2.1 Deep Blue . . . . . 354

    2.2 Virtual Reality . . . . . 354

    2.3 The Humble Earthworm . . . . . 354

3 Consciousness . . . . . 355

    3.1 Feeling and Qualia . . . . . 356

4 General Intelligence . . . . . 358

    4.1 Human Intelligence . . . . . 360

5 3D Simulation and Language ..... 363  
 6 Epistemology ..... 366  
 7 Instantiation: the Heart of Consciousness ..... 367  
 8 In a Nutshell ..... 370  
 9 Real-World AI ..... 374  
     9.1 Examples and Metaphors ..... 378  
     9.2 Math and Software..... 380  
     9.3 Barcode Example ..... 380  
     9.4 Software Design ..... 383  
 10 Conclusion ..... 385  
 References ..... 386

**Levels of Organization in General Intelligence**

*Eliezer Yudkowsky*

1 Foundations of General Intelligence ..... 389  
 2 Levels of Organization in Deliberative General Intelligence..... 397  
     2.1 Concepts: An Illustration of Principles..... 397  
     2.2 Levels of Organization in Deliberation ..... 407  
     2.3 The Code Level ..... 409  
     2.4 The Modality Level ..... 416  
     2.5 The Concept Level ..... 426  
     2.6 The Thought Level ..... 444  
     2.7 The Deliberation Level ..... 461  
 3 Seed AI ..... 476  
     3.1 Advantages of Minds-in-General ..... 480  
     3.2 Recursive Self-enhancement ..... 484  
     3.3 Infrahumanity and Transhumanity: “Human-Equivalence” as  
         Anthropocentrism ..... 489  
 4 Conclusions ..... 493  
 References ..... 496

**Index** ..... 503



<http://www.springer.com/978-3-540-23733-4>

Artificial General Intelligence

Goertzel, B.; Pennachin, C. (Eds.)

2007, XVI, 509 p., Hardcover

ISBN: 978-3-540-23733-4