Atlantis Thinking Machines

Volume 5

Series editor

Kai-Uwe Kühnberger, Osnabrück, Germany

For further volumes:
http://www.atlantis-press.com

Aims and Scope of the Series

This series publishes books resulting from theoretical research on and reproductions of general Artificial Intelligence (AI). The book series focuses on the establishment of new theories and paradigms in AI. At the same time, the series aims at exploring multiple scientific angles and methodologies, including results from research in cognitive science, neuroscience, theoretical and experimental AI, biology and from innovative interdisciplinary methodologies.

For more information on this series and our other book series, please visit our website at: www.atlantis-press.com/publications/books

AMSTERDAM—PARIS—BEIJING
ATLANTIS PRESS
Atlantis Press
29, avenue Laumière
75019 Paris, France

Ben Goertzel · Cassio Pennachin
Nil Geisweiller

Engineering General Intelligence, Part 1

A Path to Advanced AGI via Embodied Learning and Cognitive Synergy

With contributions by the OpenCog Team

Ben Goertzel
G/F 51C Lung Mei Village
Tai Po
Hong Kong
People's Republic of China

Cassio Pennachin
Igenesis
Belo Horizonte, Minas Gerais
Brazil

Nil Geisweiller
Samokov
Bulgaria

ISSN 1877-3273
ISBN 978-94-6239-026-3 ISBN 978-94-6239-027-0 (eBook)
DOI 10.2991/978-94-6239-027-0

Library of Congress Control Number: 2013953280

Preface

This is a large, two-part book with an even larger goal: To outline a practical approach to engineering software systems with general intelligence at the human level and ultimately beyond. Machines with flexible problem-solving ability, open-ended learning capability, creativity, and eventually their own kind of genius.

Part 1 of the book (Volume 5 in the Atlantis Thinking Machines book series), reviews various critical conceptual issues related to the nature of intelligence and mind. It then sketches the broad outlines of a novel, integrative architecture for Artificial General Intelligence (AGI) called CogPrime… and describes an approach for giving a young AGI system (CogPrime or otherwise) appropriate experience, so that it can develop its own smarts, creativity, and wisdom through its own experience. Along the way a formal theory of general intelligence is sketched, and a broad roadmap leading from here to human-level artificial intelligence. Hints are also given regarding how to eventually, potentially create machines advancing beyond human level—including some frankly futuristic speculations about strongly self-modifying AGI architectures with flexibility far exceeding that of the human brain.

Part 2 of the book (Volume 6 in the Atlantis Thinking Machines book series), then digs far deeper into the details of CogPrime's multiple structures, processes, and functions, culminating in a general argument as to why we believe CogPrime will be able to achieve general intelligence at the level of the smartest humans (and potentially greater), and a detailed discussion of how a CogPrime-powered virtual agent or robot would handle some simple practical tasks such as social play with blocks in a preschool context. It first describes the CogPrime software architecture and knowledge representation in detail; then reviews the cognitive cycle via which CogPrime perceives and acts in the world and reflects on itself; and next turns to various forms of learning: procedural, declarative (e.g., inference), simulative, and integrative. Methods of enabling natural language functionality in CogPrime are then discussed; and then the volume concludes with a chapter summarizing the argument that CogPrime can lead to human-level (and eventually perhaps greater) AGI, and a chapter giving a thought experiment describing the internal dynamics via which a completed CogPrime system might solve the problem of obeying the request "Build me something with blocks that I haven't seen before."

The chapters here are written to be read in linear order—and if consumed thus, they tell a coherent story about how to get from here to advanced AGI.

However, we suggest the impatient reader may wish to take a quick look at the final chapter of Part 2, after reading Chaps. 1–3 of Part 1. This final chapter gives a broad overview of why we think the CogPrime design will work, in a way that depends on the technical details of the previous chapters, but (we believe) not so sensitively as to be incomprehensible without them.

This is admittedly an unusual sort of book, mixing demonstrated conclusions with unproved conjectures in a complex way, all oriented toward an extraordinarily ambitious goal. Further, the chapters are somewhat variant in their levels of detail—some very nitty-gritty, some more high level, with much of the variation due to how much concrete work has been done on the topic of the chapter at time of writing. However, it is important to understand that the ideas presented here are not mere armchair speculation—they are currently being used as the basis for an open-source software project called OpenCog, which is being worked on by software developers around the world. Right now OpenCog embodies only a percentage of the overall CogPrime design as described here. But if OpenCog continues to attract sufficient funding or volunteer interest, then the ideas presented in these volumes will be validated or refuted via practice (As a related note: here and there in this book, we will refer to the "current" CogPrime implementation (in the OpenCog framework); in all cases this refers to OpenCog as of late 2013).

To state one believes one knows a workable path to creating a human-level (and potentially greater) general intelligence is to make a dramatic statement, given the conventional way of thinking about the topic in the contemporary scientific community. However, we feel that once a little more time has passed, the topic will lose its drama (if not its interest and importance), and it will be widely accepted that there are *many* ways to create intelligent machines—some simpler and some more complicated; some more brain-like or human-like and some less so; some more efficient and some more wasteful of resources; etc. We have little doubt that, from the perspective of AGI science 50 or 100 years hence (and probably even 10–20 years hence), the specific designs presented here will seem awkward, messy, inefficient, and circuitous in various respects. But that is how science and engineering progress. Given the current state of knowledge and understanding, having any concrete, comprehensive design, and plan for creating AGI is a significant step forward; and it is in this spirit that we present here our thinking about the CogPrime architecture and the nature of general intelligence.

In the words of Sir Edmund Hillary, the first to scale Everest: "Nothing Venture, Nothing Win."

Prehistory of the Book

The writing of this book began in earnest in 2001, at which point it was informally referred to as "The Novamente Book." The original "Novamente Book" manuscript ultimately got too big for its own britches, and subdivided into a number of different works—*The Hidden Pattern* (Goertzel 2006), a philosophy of mind book

published in 2006; *Probabilistic Logic Networks* (Goertzel et al. 2008), a more technical work published in 2008; *Real World Reasoning* (Goertzel et al. 2011), a sequel to *Probabilistic Logic Networks* published in 2011; and the two parts of this book.

The ideas described in this book have been the collaborative creation of multiple overlapping communities of people over a long period of time. The vast bulk of the writing here was done by Ben Goertzel; but Cassio Pennachin and Nil Geisweiller made sufficient writing, thinking, and editing contributions over the years to more than merit their inclusion as coauthors. Further, many of the chapters here have coauthors beyond the three main coauthors of the book; and the set of chapter coauthors does not exhaust the set of significant contributors to the ideas presented.

The core concepts of the CogPrime design and the underlying theory were conceived by Ben Goertzel in the period 1995–1996 when he was a Research Fellow at the University of Western Australia; but those early ideas have been elaborated and improved by many more people than can be listed here (as well as by Ben's ongoing thinking and research). The collaborative design process ultimately resulting in CogPrime started in 1997 when Intelligenesis Corp. was formed—the Webmind AI Engine created in Intelligenesis's research group during 1997–2001 was the predecessor to the Novamente Cognition Engine created at Novamente LLC during 2001–2008, which was the predecessor to CogPrime.

Online Appendices

Just one more thing before getting started! This book originally had even more chapters than the ones currently presented in Parts 1 and 2. In order to decrease length and increase focus, however, a number of chapters dealing with peripheral—yet still relevant and interesting—matters were moved to online appendices. These may be downloaded in a single PDF file at http://goertzel.org/engineering_general_Intelligence_appendices_B-H.pdf. The titles of these appendices are:

- Appendix A: Possible Worlds Semantics and Experiential Semantics
- Appendix B: Steps Toward a Formal Theory of Cognitive Structure and Dynamics
- Appendix C: Emergent Reflexive Mental Structures
- Appendix D: GOLEM: Toward an AGI Meta-Architecture Enabling Both Goal Preservation and Radical Self-Improvement
- Appendix E: Lojban++: A Novel Linguistic Mechanism for Teaching AGI Systems
- Appendix F: Possible Worlds Semantics and Experiential Semantics
- Appendix G: PLN and the Brain
- Appendix H: Propositions About Environments in Which CogPrime Components Are Useful

None of these are critical to understanding the key ideas in the book, which is why they were relegated to online appendices. However, reading them will deepen your understanding of the conceptual and formal perspectives underlying the CogPrime design. These appendices are referred to here and there in the text of the main book.

References

B. Goertzel, *The Hidden Pattern* (Brown Walker, Boca Raton, 2006)
B. Goertzel, M. Ikle, I. Goertzel, A. Heljakka, *Probabilistic Logic Networks* (Springer, Heidelberg, 2008)
B. Goertzel, N. Geisweiller, L. Coelho, P. Janicic, C. Pennachin, *Real World Reasoning* (Atlantis, Oxford, 2011)

Acknowledgments

For sake of simplicity, this acknowledgments section is presented from the perspective of the primary author, Ben Goertzel. Ben will thus begin by expressing his thanks to his primary coauthors, Cassio Pennachin (Collaborator since 1998) and Nil Geisweiller (Collaborator since 2005). Without outstandingly insightful, deep-thinking colleagues like you, the ideas presented here—let alone the book itself—would not have developed nearly as effectively as what has happened. Similar thanks also go to the other OpenCog Collaborators who have coauthored various chapters of the book.

Beyond the coauthors, huge gratitude must also be extended to everyone who has been involved with the OpenCog project, and/or was involved in Novamente LLC and Webmind Inc. before that. We are grateful to all of you for your collaboration and intellectual companionship!

Building a thinking machine is a huge project, too big for any one human; it will take a team and I'm happy to be part of a great one. It is through the genius of human collectives, going beyond any individual human mind, that genius machines are going to be created.

A tiny, incomplete sample from the long list of those others deserving thanks is:

- Ken Silverman and Gwendalin Qi Aranya (formerly Gwen Goertzel), both of whom listened to me talk at inordinate length about many of the ideas presented here a long, long time before anyone else was interested in listening. Ken and I schemed some AGI designs at Simon's Rock College in 1983, years before we worked together on the Webmind AI Engine.
- Allan Combs, who got me thinking about consciousness in various different ways, at a very early point in my career. I'm very pleased to still count Allan as a friend and sometime Collaborator! Fred Abraham as well, for introducing me to the intersection of chaos theory and cognition, with a wonderful flair. George Christos, a deep AI/math/physics thinker from Perth, for reawakening my interest in attractor neural nets and their cognitive implications, in the mid-1990s.
- All of the 130 staff of Webmind Inc. during 1998–2001 while that remarkable, ambitious, peculiar AGI-oriented firm existed. Special shout-outs to the "Voice of Reason" Pei Wang and the "Siberian Madmind" Anton Kolonin, Mike Ross, Cate Hartley, Karin Verspoor, and the tragically prematurely deceased Jeff

Pressing (compared to whom we are all mental midgets), who all made serious conceptual contributions to my thinking about AGI. Lisa Pazer and Andy Siciliano who made Webmind happen on the business side. And of course Cassio Pennachin, a coauthor of this book; and Ken Silverman, who coarchitected the whole Webmind system and vision with me from the start.

- The Webmind Diehards, who helped begin the Novamente project that succeeded Webmind beginning in 2001: Cassio Pennachin, Stephan Vladimir Bugaj, Takuo Henmi, Matthew Ikle', Thiago Maia, Andre Senna, Guilherme Lamacie, and Saulo Pinto.
- Those who helped get the Novamente project off the ground and keep it progressing over the years, including some of the Webmind Diehards and also Moshe Looks, Bruce Klein, Izabela Lyon Freire, Chris Poulin, Murilo Queiroz, Predrag Janicic, David Hart, Ari Heljakka, Hugo Pinto, Deborah Duong, Paul Prueitt, Glenn Tarbox, Nil Geisweiller and Cassio Pennachin (the coauthors of this book), Sibley Verbeck, Jeff Reed, Pejman Makhfi, Welter Silva, Lukasz Kaiser, and more.
- Izabela Lyon Freire, for endless wide-ranging discussions on AI, philosophy of mind and related topics, which helped concretize many of the ideas here.
- David Hart, for his various, inspired IT and editing help over the years, including lots of editing and formatting help on an earlier version of this book which took the form of a Wikibook.
- A group of volunteers who helped proofread a draft of this manuscript, including Brett Woodward, Terry Ward, Dario Garcia Gasulla, Tom Heiman as well as others. Tim Josling's efforts in this regard were especially thorough and helpful.
- All those who have helped with the OpenCog system, including Linas Vepstas, Joel Pitt, Jared Wigmore/Jade O'Neill, Zhenhua Cai, Deheng Huang, Shujing Ke, Lake Watkins, Alex van der Peet, Samir Araujo, Fabricio Silva, Yang Ye, Shuo Chen, Michel Drenthe, Ted Sanders, Gustavo Gama and of course Nil, and Cassio again. Tyler Emerson and Eliezer Yudkowsky, for choosing to have the Singularity Institute for AI (now MIRI) provide seed funding for OpenCog.
- The numerous members of the AGI community who have tossed around AGI ideas with me since the first AGI conference in 2006, including but definitely not limited to: Stan Franklin, Juergen Schmidhuber, Marcus Hutter, Kai-Uwe Küehnberger, Stephen Reed, Blerim Enruli, Kristinn Thorisson, Joscha Bach, Abram Demski, Itamar Arel, Mark Waser, Randal Koene, Paul Rosenbloom, Zhongzhi Shi, Steve Omohundro, Bill Hibbard, Eray Ozkural, Brandon Rohrer, Ben Johnston, John Laird, Shane Legg, Selmer Bringsjord, Anders Sandberg, Alexei Samsonovich, Wlodek Duch, and more.
- The inimitable "Artilect Warrior" Hugo de Garis, who (when he was working at Xiamen University) got me started working on AGI in the Orient (and introduced me to my wife Ruiting in the process). And Changle Zhou, who brought Hugo to Xiamen and generously shared his brilliant research students with Hugo and me. And Min Jiang, Collaborator of Hugo and Changle, a deep AGI thinker who is helping with OpenCog theory and practice at time of writing.

- Gino Yu, who got me started working on AGI here in Hong Kong, where I am living at time of writing. As of 2013 the bulk of OpenCog work is occurring in Hong Kong via a research grant that Gino and I obtained together.
- Dan Stoicescu, whose funding helped Novamente through some tough times.
- Jeffrey Epstein, whose visionary funding of my AGI research has helped me through a number of tight spots over the years. At time of writing, Jeffrey is helping support the OpenCog Hong Kong project.
- Zeger Karssen, Founder of Atlantis Press, who conceived the *Thinking Machines* book series in which this book appears, and who has been a strong supporter of the AGI conference series from the beginning.
- My wonderful wife Ruiting Lian, a source of fantastic amounts of positive energy for me since we became involved 4 years ago. Ruiting has listened to me discuss the ideas contained here time and time again, often with judicious and insightful feedback (as she is an excellent AI Researcher in her own right); and has been wonderfully tolerant of me diverting numerous evenings and weekends to getting this book finished (as well as to other AGI-related pursuits). And my parents Ted and Carol and kids Zar, Zeb, and Zade, who have also indulged me in discussions on many of the themes discussed here on countless occasions! And my dear, departed grandfather Leo Zwell, for getting me started in science.
- Crunchkin and Pumpkin, for regularly getting me away from the desk to stroll around the village where we live; many of my best ideas about AGI and other topics have emerged while walking with my furry four-legged family members.

September 2013 Ben Goertzel

Contents

Acronyms

AA	Attention Allocation
ADF	Automatically Defined Function (in the context of Genetic Programming)
AF	Attentional Focus
AGI	Artificial General Intelligence
AV	Attention Value
BD	Behavior Description
C-space	Configuration Space
CBV	Coherent Blended Volition
CEV	Coherent Extrapolated Volition
CGGP	Contextually Guided Greedy Parsing
CSDLN	Compositional Spatiotemporal Deep Learning Network
CT	Combo Tree
ECAN	Economic Attention Network
ECP	Embodied Communication Prior
EPW	Experiential Possible Worlds (semantics)
FCA	Formal Concept Analysis
FI	Fisher Information
FIM	Frequent Itemset Mining
FOI	First Order Inference
FOPL	First Order Predicate Logic
FOPLN	First Order PLN
FS-MOSES	Feature Selection MOSES (i.e. MOSES with feature selection integrated a la LIFES)
GA	Genetic Algorithms
GB	Global Brain
GEOP	Goal Evaluator Operating Procedure (in a GOLEM context)
GIS	Geospatial Information System
GOLEM	Goal-Oriented LEarning Meta-architecture
GP	Genetic Programming
HOI	Higher-Order Inference
HOPLN	Higher-Order PLN

HR	Historical Repository (in a GOLEM context)
HTM	Hierarchical Temporal Memory
IA	(Allen) Interval Algebra (an algebra of temporal intervals)
IRC	Imitation/Reinforcement/Correction (Learning)
LIFES	Learning-Integrated Feature Selection
LTI	Long Term Importance
MA	Mind Agent
MOSES	Meta-Optimizing Semantic Evolutionary Search
MSH	Mirror System Hypothesis
NARS	Non-Axiomatic Reasoning System
NLGen	A specific software component within OpenCog, which provides one way of dealing with Natural Language Generation
OCP	OpenCogPrime
OP	Operating Program (in a GOLEM context)
PEPL	Probabilistic Evolutionary Procedure Learning (e.g. MOSES)
PLN	Probabilistic Logic Networks
RCC	Region Connection Calculus
RelEx	A specific software component within OpenCog, which provides one way of dealing with natural language Relationship Extraction
SAT	Boolean SATisfaction, as a mathematical/computational problem
SMEPH	Self-Modifying Evolving Probabilistic Hypergraph
SRAM	Simple Realistic Agents Model
STI	Short Term Importance
STV	Simple Truth Value
TV	Truth Value
VLTI	Very Long Term Importances
WSPS	Whole-Sentence Purely-Syntactic Parsing