

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)

Engineering General Intelligence, Part 1

A Path to Advanced AGI via Embodied Learning and
Cognitive Synergy

Goertzel, B.; Pennachin, C.; Geisweiller, N.

2014, XXII, 409 p. 105 illus., 32 illus. in color., Hardcover

ISBN: 978-94-6239-026-3

A product of Atlantis Press