Contents

1 Introduction ........................................ 1
  1.1 AI Returns to Its Roots ........................... 1
  1.2 AGI Versus Narrow AI ............................ 2
  1.3 CogPrime ....................................... 3
  1.4 The Secret Sauce ................................. 4
  1.5 Extraordinary Proof? ............................ 5
  1.6 Potential Approaches to AGI ...................... 6
    1.6.1 Build AGI from Narrow AI ..................... 6
    1.6.2 Enhancing Chatbots ............................ 7
    1.6.3 Emulating the Brain ........................... 7
    1.6.4 Evolve an AGI ................................ 8
    1.6.5 Derive an AGI Design Mathematically ........ 8
    1.6.6 Use Heuristic Computer Science Methods ...... 9
    1.6.7 Integrative Cognitive Architecture ............ 9
    1.6.8 Can Digital Computers Really Be Intelligent? . 10
  1.7 Five Key Words .................................. 11
    1.7.1 Memory and Cognition in CogPrime .......... 11
  1.8 Virtually and Robotically Embodied AI ............ 13
  1.9 Language Learning ................................ 13
  1.10 AGI Ethics ..................................... 14
  1.11 Structure of the Book ............................ 15
  1.12 Key Claims of the Book .......................... 15

Part I Overview of the CogPrime Architecture

2 A Brief Overview of CogPrime .......................... 21
  2.1 Introduction ..................................... 21
  2.2 High-Level Architecture of CogPrime ............... 21
  2.3 Current and Prior Applications of OpenCog ........ 23
    2.3.1 Transitioning from Virtual Agents to a Physical Robot ....................... 25
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4</td>
<td>Memory Types and Associated Cognitive Processes</td>
<td>30</td>
</tr>
<tr>
<td>2.4.1</td>
<td>Cognitive Synergy in PLN</td>
<td>32</td>
</tr>
<tr>
<td>2.5</td>
<td>Goal-Oriented Dynamics in CogPrime</td>
<td>35</td>
</tr>
<tr>
<td>2.6</td>
<td>Analysis and Synthesis Processes in CogPrime</td>
<td>36</td>
</tr>
<tr>
<td>2.7</td>
<td>Conclusion</td>
<td>40</td>
</tr>
<tr>
<td>3</td>
<td>Build Me Something I Haven't Seen: A CogPrime Thought Experiment</td>
<td>41</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>41</td>
</tr>
<tr>
<td>3.2</td>
<td>Roles of Selected Cognitive Processes</td>
<td>42</td>
</tr>
<tr>
<td>3.3</td>
<td>A Semi-Narrative Treatment</td>
<td>54</td>
</tr>
<tr>
<td>3.4</td>
<td>Conclusion</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>Part II Artificial and Natural General Intelligence</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>What is Human-Like General Intelligence?</td>
<td>61</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>61</td>
</tr>
<tr>
<td>4.1.1</td>
<td>What is General Intelligence?</td>
<td>62</td>
</tr>
<tr>
<td>4.1.2</td>
<td>What is Human-Like General Intelligence</td>
<td>62</td>
</tr>
<tr>
<td>4.2</td>
<td>Commonly Recognized Aspects of Human-Like Intelligence</td>
<td>63</td>
</tr>
<tr>
<td>4.3</td>
<td>Further Characterizations of Human-Like Intelligence</td>
<td>66</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Competencies Characterizing Human-Like Intelligence</td>
<td>66</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Gardner’s Theory of Multiple Intelligences</td>
<td>67</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Newell’s Criteria for a Human Cognitive Architecture</td>
<td>68</td>
</tr>
<tr>
<td>4.3.4</td>
<td>Intelligence and Creativity</td>
<td>69</td>
</tr>
<tr>
<td>4.4</td>
<td>Preschool as a View into Human-Like General Intelligence</td>
<td>70</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Design for an AGI Preschool</td>
<td>72</td>
</tr>
<tr>
<td>4.5</td>
<td>Integrative and Synergetic Approaches to Artificial General Intelligence</td>
<td>74</td>
</tr>
<tr>
<td>4.5.1</td>
<td>Achieving Human-Like Intelligence via Cognitive Synergy</td>
<td>75</td>
</tr>
<tr>
<td>5</td>
<td>A Patternist Philosophy of Mind</td>
<td>77</td>
</tr>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>77</td>
</tr>
<tr>
<td>5.2</td>
<td>Some Patternist Principles</td>
<td>78</td>
</tr>
<tr>
<td>5.3</td>
<td>Cognitive Synergy</td>
<td>83</td>
</tr>
</tbody>
</table>
5.4 The General Structure of Cognitive Dynamics:
Analysis and Synthesis ........................................ 85
5.4.1 Component-Systems and Self-Generating
Systems .......................................................... 85
5.4.2 Analysis and Synthesis ..................................... 87
5.4.3 The Dynamic of Iterative Analysis
and Synthesis ...................................................... 90
5.4.4 Self and Focused Attention as Approximate
Attractors of the Dynamic of Iterated
Forward-Analysis ................................................. 91
5.4.5 Conclusion .................................................... 94
5.5 Perspectives on Machine Consciousness .................. 95
5.6 Postscript: Formalizing Pattern ............................ 96

6 Brief Survey of Cognitive Architectures ...................... 101
6.1 Introduction ..................................................... 101
6.2 Symbolic Cognitive Architectures ......................... 103
6.2.1 SOAR ........................................................ 104
6.2.2 ACT-R ....................................................... 105
6.2.3 Cyc and Texai ............................................... 107
6.2.4 NARS ........................................................ 107
6.2.5 GLAIR and SNePS ....................................... 108
6.3 Emergentist Cognitive Architectures ....................... 109
6.3.1 DeSTIN: A Deep Reinforcement Learning
Approach to AGI .................................................. 111
6.3.2 Developmental Robotics Architectures ................ 117
6.4 Hybrid Cognitive Architectures ............................. 119
6.4.1 Neural Versus Symbolic; Global Versus Local .......... 121
6.5 Globalist Versus Localist Representations ............... 124
6.5.1 CLARION .................................................... 125
6.5.2 The Society of Mind and the Emotion Machine ....... 126
6.5.3 DUAL ......................................................... 126
6.5.4 4D/RCS ...................................................... 128
6.5.5 PolyScheme ................................................. 130
6.5.6 Joshua Blue ............................................... 130
6.5.7 LIDA ......................................................... 131
6.5.8 The Global Workspace .................................... 131
6.5.9 The LIDA Cognitive Cycle ............................... 132
6.5.10 Psi and MicroPsi ......................................... 136
6.5.11 The Emergence of Emotion in the Psi Model ...... 139
6.5.12 Knowledge Representation, Action Selection
and Planning in Psi ............................................... 140
6.5.13 Psi Versus CogPrime ..................................... 142
7  A Generic Architecture of Human-Like Cognition ............................. 143
  7.1  Introduction .......................................................... 143
  7.2  Key Ingredients of the Integrative Human-Like
    Cognitive Architecture Diagram ....................................... 144
  7.3  An Architecture Diagram for Human-Like
    General Intelligence .................................................... 146
  7.4  Interpretation and Application of the Integrative Diagram .......... 152

Part III  Toward a General Theory of General Intelligence

8  A Formal Model of Intelligent Agents ........................................ 157
  8.1  Introduction .......................................................... 157
  8.2  A Simple Formal Agents Model (SRAM) ................................ 158
      8.2.1  Goals ............................................................ 159
      8.2.2  Memory Stores .................................................. 160
      8.2.3  The Cognitive Schematic ...................................... 162
  8.3  Toward a Formal Characterization of Real-World
    General Intelligence .................................................... 163
      8.3.1  Biased Universal Intelligence ................................. 165
      8.3.2  Connecting Legg and Hutter’s Model of
            Intelligent Agents to the Real World ....................... 166
      8.3.3  Pragmatic General Intelligence ............................... 167
      8.3.4  Incorporating Computational Cost ............................. 168
      8.3.5  Assessing the Intelligence of Real-World Agents ......... 169
  8.4  Intellectual Breadth: Quantifying the Generality of
      an Agent’s Intelligence ............................................... 171
  8.5  Conclusion .......................................................... 172

9  Cognitive Synergy .................................................................... 173
  9.1  Introduction .......................................................... 173
  9.2  Cognitive Synergy ...................................................... 174
  9.3  Cognitive Synergy in CogPrime ...................................... 177
      9.3.1  Cognitive Processes in CogPrime ............................ 177
  9.4  Some Critical Synergies ................................................. 182
  9.5  Cognitive Synergy for Procedural
      and Declarative Learning .............................................. 182
      9.5.1  Cognitive Synergy in MOSES ................................. 187
      9.5.2  Cognitive Synergy in PLN ...................................... 190
  9.6  Is Cognitive Synergy Tricky? ......................................... 191
      9.6.1  The Puzzle: Why Is It So Hard to Measure
            Partial Progress Toward Human-Level AGI? ............... 192
      9.6.2  A Possible Answer: Cognitive Synergy Is Tricky! ....... 193
      9.6.3  Conclusion ....................................................... 194
10 General Intelligence in the Everyday Human World 197
  10.1 Introduction ......................................................... 197
  10.2 Some Broad Properties of the Everyday World that Help Structure Intelligence ........................................... 198
  10.3 Embodied Communication ........................................ 199
    10.3.1 Generalizing the Embodied Communication Prior ................................................................. 202
  10.4 Naive Physics .......................................................... 203
    10.4.1 Objects, Natural Units and Natural Kinds ...................... 204
    10.4.2 Events, Processes and Causality ............................... 204
    10.4.3 Stuff, States of Matter, Qualities ............................ 205
    10.4.4 Surfaces, Limits, Boundaries, Media ........................ 205
    10.4.5 What Kind of Physics is Needed to Foster Human-Like Intelligence? ...................................... 206
  10.5 Folk Psychology ..................................................... 208
    10.5.1 Motivation, Requiredness, Value .............................. 208
  10.6 Body and Mind ...................................................... 208
    10.6.1 The Human Sensorium ........................................... 209
    10.6.2 The Human Body’s Multiple Intelligences .................... 210
  10.7 The Extended Mind and Body ....................................... 213
  10.8 Conclusion ............................................................ 213

11 A Mind-World Correspondence Principle 215
  11.1 Introduction ........................................................... 215
  11.2 What Might a General Theory of General Intelligence Look Like? ......................................................... 216
  11.3 Steps Toward A (Formal) General Theory of General Intelligence ........................................................... 217
  11.4 The Mind-World Correspondence Principle ....................... 219
  11.5 How Might the Mind-World Correspondence Principle Be Useful? .............................................................. 220
  11.6 Conclusion ............................................................. 221

Part IV Cognitive and Ethical Development

12 Stages of Cognitive Development 225
  12.1 Introduction ........................................................... 225
  12.2 Piagetan Stages in the Context of a General Systems Theory of Development ........................................... 226
  12.3 Piaget’s Theory of Cognitive Development ....................... 227
    12.3.1 Perry’s Stages .................................................... 231
    12.3.2 Keeping Continuity in Mind .................................... 232
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.4</td>
<td>Piaget’s Stages in the Context of Uncertain Inference</td>
<td>232</td>
</tr>
<tr>
<td>12.4.1</td>
<td>The Infantile Stage</td>
<td>234</td>
</tr>
<tr>
<td>12.4.2</td>
<td>The Concrete Stage</td>
<td>236</td>
</tr>
<tr>
<td>12.4.3</td>
<td>The Formal Stage</td>
<td>240</td>
</tr>
<tr>
<td>12.4.4</td>
<td>The Reflexive Stage</td>
<td>242</td>
</tr>
<tr>
<td>13</td>
<td>The Engineering and Development of Ethics</td>
<td>245</td>
</tr>
<tr>
<td>13.1</td>
<td>Introduction</td>
<td>245</td>
</tr>
<tr>
<td>13.2</td>
<td>Review of Current Thinking on the Risks of AGI</td>
<td>246</td>
</tr>
<tr>
<td>13.3</td>
<td>The Value of an Explicit Goal System</td>
<td>249</td>
</tr>
<tr>
<td>13.4</td>
<td>Ethical Synergy</td>
<td>251</td>
</tr>
<tr>
<td>13.4.1</td>
<td>Stages of Development of Declarative Ethics</td>
<td>252</td>
</tr>
<tr>
<td>13.4.2</td>
<td>Stages of Development of Empathic Ethics</td>
<td>255</td>
</tr>
<tr>
<td>13.4.3</td>
<td>An Integrative Approach to Ethical Development</td>
<td>257</td>
</tr>
<tr>
<td>13.4.4</td>
<td>Integrative Ethics and Integrative AGI</td>
<td>259</td>
</tr>
<tr>
<td>13.5</td>
<td>Clarifying the Ethics of Justice: Extending the Golden Rule in to a Multifactorial Ethical Model</td>
<td>261</td>
</tr>
<tr>
<td>13.5.1</td>
<td>The Golden Rule and the Stages of Ethical Development</td>
<td>264</td>
</tr>
<tr>
<td>13.5.2</td>
<td>The Need for Context-Sensitivity and Adaptiveness in Deploying Ethical Principles</td>
<td>266</td>
</tr>
<tr>
<td>13.6</td>
<td>The Ethical Treatment of AGIs</td>
<td>269</td>
</tr>
<tr>
<td>13.6.1</td>
<td>Possible Consequences of Depriving AGIs of Freedom</td>
<td>271</td>
</tr>
<tr>
<td>13.6.2</td>
<td>AGI Ethics as Boundaries Between Humans and AGIs Become Blurred</td>
<td>272</td>
</tr>
<tr>
<td>13.7</td>
<td>Possible Benefits of Closely Linking AGIs to the Global Brain</td>
<td>273</td>
</tr>
<tr>
<td>13.7.1</td>
<td>The Importance of Fostering Deep, Consensus-Building Interactions Between People with Divergent Views</td>
<td>275</td>
</tr>
<tr>
<td>13.8</td>
<td>Possible Benefits of Creating Societies of AGIs</td>
<td>277</td>
</tr>
<tr>
<td>13.9</td>
<td>AGI Ethics as Related to Various Future Scenarios</td>
<td>278</td>
</tr>
<tr>
<td>13.9.1</td>
<td>Capped Intelligence Scenarios</td>
<td>278</td>
</tr>
<tr>
<td>13.9.2</td>
<td>Superintelligent AI: Soft-Takeoff Scenarios</td>
<td>279</td>
</tr>
<tr>
<td>13.9.3</td>
<td>Superintelligent AI: Hard-Takeoff Scenarios</td>
<td>280</td>
</tr>
<tr>
<td>13.9.4</td>
<td>Global Brain Mindplex Scenarios</td>
<td>281</td>
</tr>
<tr>
<td>13.10</td>
<td>Conclusion: Eight Ways to Bias AGI Toward Friendliness</td>
<td>284</td>
</tr>
<tr>
<td>13.10.1</td>
<td>Encourage Measured Co-Advancement of AGI Software and AGI Ethics Theory</td>
<td>286</td>
</tr>
<tr>
<td>13.10.2</td>
<td>Develop Advanced AGI Sooner Not Later</td>
<td>286</td>
</tr>
</tbody>
</table>
Part V  Networks for Explicit and Implicit Knowledge Representation

14 Local, Global and Glocal Knowledge Representation ................. 291
   14.1 Introduction ................................................. 291
   14.2 Localized Knowledge Representation Using Weighted, Labeled Hypergraphs ........................................ 292
       14.2.1 Weighted, Labeled Hypergraphs ................. 292
   14.3 Atoms: Their Types and Weights ...................................... 293
       14.3.1 Some Basic Atom Types ......................... 294
       14.3.2 Variable Atoms ........................................... 296
       14.3.3 Logical Links .................................... 297
       14.3.4 Temporal Links .................................... 299
       14.3.5 Associative Links ................................ 300
       14.3.6 Procedure Nodes .................................. 301
       14.3.7 Links for Special External Data Types .......... 302
       14.3.8 Truth Values and Attention Values ............... 302
   14.4 Knowledge Representation via Attractor Neural Networks ....... 303
       14.4.1 The Hopfield Neural Net Model .................. 303
       14.4.2 Knowledge Representation via Cell Assemblies .... 304
   14.5 Neural Foundations of Learning .................................. 305
       14.5.1 Hebbian Learning ................................... 305
       14.5.2 Virtual Synapses and Hebbian Learning Between Assemblies ........................................ 306
       14.5.3 Neural Darwinism ................................... 307
   14.6 Glocal Memory ............................................... 308
       14.6.1 A Semi-Formal Model of Glocal Memory ........... 310
       14.6.2 Glocal Memory in the Brain ....................... 312
       14.6.3 Glocal Hopfield Networks ........................... 315
       14.6.4 Neural-Symbolic Glocality in CogPrime .......... 317

15 Representing Implicit Knowledge via Hypergraphs .................... 319
   15.1 Introduction ................................................. 319
   15.2 Key Vertex and Edge Types ..................................... 320
   15.3 Derived Hypergraphs ........................................... 320
       15.3.1 SMEPH Vertices .................................. 321
       15.3.2 SMEPH Edges .................................... 322
   15.4 Implications of Patternist Philosophy for Derived Hypergraphs of Intelligent Systems .......................... 323
       15.4.1 SMEPH Principles in CogPrime ................. 324
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
A product of Atlantis Press