PREPRINT

Towards Artificial General Intelligence (AGI): Architectural Frameworks, Mathematical Formalisms, and Hardware Enablement

Raymund K.D. Kho, et al., The Center for Research and Development Hong Kong (CRD-HK) Tsuen Wan N.T. - Hong Kong, dept. Micro-Electronics, 1 June 2025, Spectral Sensing, info@crd-hk.com, remykho@pm.me. DOI: 10.17613/et1p0-0e714

Abstract

The development of Artificial General Intelligence (AGI), characterized by human-like cognitive adaptability and broad problem-solving capabilities, represents a formidable challenge within computer science. While Artificial Narrow Intelligence (ANI) has demonstrated considerable success in domain-specific applications, it fundamentally lacks the robust and flexible reasoning requisite for generalized problem-solving. This paper presents a comprehensive survey of the architectural frameworks and mathematical formalisms underpinning contemporary AGI research. Furthermore, it critically examines the requisite hardware platforms and optimization strategies essential for the practical realization of these computationally intensive models.

This review commences with an exploration of the philosophical underpinnings and core definitions of intelligence. Subsequently, a detailed analysis of prominent cognitive architectures is presented, encompassing established symbolic systems such as SOAR and ACT-R, alongside emergent neural-symbolic hybrid approaches. The paper then proceeds to examine key learning paradigms, including reinforcement learning, self-supervised learning, and meta-learning, elucidating their respective mathematical models and the practical techniques employed to facilitate their training. Moreover, crucial topics in knowledge representation are explored, ranging from formal logics and ontologies to probabilistic and causal models. A dedicated section is devoted to the examination of how high-performance hardware and sophisticated software optimizations collectively provide the necessary computational substrate for the development and deployment of AGI concepts.

The central argument advanced is that substantive progress toward AGI necessitates a synergistic integration of these diverse approaches. This involves combining the perceptual prowess inherent in connectionist models with the explicit reasoning capabilities of symbolic systems, all of which must be accelerated by advanced, heterogeneous hardware and efficient algorithms. The paper concludes by summarizing the primary challenges confronting AGI research, including the common sense problem and the critical issue of AI alignment, while also outlining promising directions for future inquiry where the co-evolution of software and hardware is paramount.

1. Introduction

1.1. Defining Artificial General Intelligence (AGI)

The field of artificial intelligence (AI) has witnessed remarkable progress in recent decades, primarily through the development of Artificial Narrow Intelligence (ANI). These specialized systems have achieved superhuman performance in specific domains, ranging from game playing to medical diagnosis. However, ANI systems fundamentally lack the versatility and comprehensive understanding characteristic of human intellect. Artificial General Intelligence (AGI) represents a paradigm shift beyond these narrow applications, aiming to develop machines possessing a broad, adaptable, and autonomous intellectual capacity comparable to that of a human. The defining characteristics of AGI extend beyond mere task proficiency. They encompass generality, allowing for effective operation across diverse domains; autonomy, enabling independent function without constant human oversight; and self-improvement, facilitating the continuous enhancement of capabilities. Crucially, AGI systems are envisioned to exhibit common sense reasoning, creativity in generating novel solutions, and a genuine understanding of complex information, moving beyond superficial pattern recognition. The ambitious nature of AGI necessitates robust and multifaceted evaluation methodologies. Unlike ANI, which can often be assessed against well-defined metrics, the comprehensive intelligence envisioned for AGI presents a significant measurement challenge. Consequently, various benchmarks have been proposed to capture different facets of general intelligence. These include the classic Turing Test [Turing, 1950], which probes a system's ability to exhibit human-like conversational intelligence; the Coffee Test [Wozniak, 2010], assessing practical autonomy in everyday environments;

the Robot College Student Test [Goertzel, 2012], evaluating learning and adaptation in a complex academic setting; and the Lovelace Test [Bringsjord et al., 2001], focusing on a system's capacity for genuine creativity and unpredictable innovation. These diverse evaluative frameworks underscore the complexity inherent in defining and measuring true general intelligence in artificial systems.

1.2. The Grand Challenge of AGI

The ambition for AGI dates back to the origins of artificial intelligence [McCarthy et al., 1955], though recent decades have seen most practical success in the realm of ANI. The development of AGI remains a profound challenge due to several core difficulties, including the "common sense problem" of encoding vast amounts of implicit human knowledge, the role of embodied cognition in learning [Pfeifer & Bongard, 2006], and the need for open-ended learning capabilities. The potential impact of achieving AGI is immense, promising to revolutionize science, reshape society, and redefine humanity's future [Bostrom, 2014].

1.3. Paper Structure and Scope

This paper surveys the theoretical frameworks and mathematical formalisms that underpin the quest for Artificial General Intelligence (AGI). It further analyzes the hardware and algorithmic techniques necessary for the realization of these theories. The focus is on conceptual structures, their mathematical underpinnings, and their computational requirements. This approach acknowledges the deeply interdisciplinary nature of AGI research, which draws from cognitive science, neuroscience, philosophy, computer science, and hardware engineering. For instance, recent advancements in autonomous systems, exemplify the complex interplay of advanced hardware, sophisticated algorithms, and real-world application that characterizes cutting-edge AI development, albeit in a narrow domain [Raymund K.D. Kho et al., 2025].

2. Foundational Concepts and Philosophical Underpinnings

2.1. Intelligence: A Multifaceted Construct

Intelligence, whether biological or artificial, is a complex construct comprising perception, reasoning, learning, problem-solving, and communication [Legg & Hutter, 2007]. The role of embodiment—the interaction with a physical environment—is increasingly seen as crucial for developing grounded understanding. While emotion is central to human intelligence, its necessity for AGI remains a topic of debate.

2.2. Philosophical Debates Relevant to AGI

Several long-standing philosophical debates are critical to AGI research. These include the distinction between Strong AI and Weak AI; the Symbol Grounding Problem, which questions how symbols acquire meaning [Harnad, 1990]; and Searle's Chinese Room Argument, which challenges the notion of machine understanding [Searle, 1980]. The Frame Problem, concerning the scope of knowledge required for an action, also remains a significant hurdle [McCarthy & Hayes, 1969].

2.3. The Problem of Generality

A central question for AGI is how an agent can learn to perform any intellectual task. Research into transfer learning [Pan & Yang, 2010], multi-task learning, and lifelong learning represents incremental steps toward this goal [Thrun & Pratt, 2012]. The primary challenge lies in creating systems that can adapt to novel situations and manage unknown unknowns.

3. Cognitive Architectures for AGI

3.1. Overview of Cognitive Architectures

Cognitive architectures provide integrated frameworks for intelligent behaviour. They can be broadly categorized into symbolic, connectionist, and hybrid approaches, each with distinct strengths and weaknesses [Newell, 1990].

3.2. Symbolic Cognitive Architectures

- **SOAR (State, Operator, And Result):** This architecture is based on production rules (If-Then statements) that operate within defined problem spaces. Its primary learning mechanism is "chunking," where successful chains of operations are consolidated into new rules [Laird, 2012].
    - **Formalism:** Given a set of states S and operators O, a result function is defined as $R: S \times O \rightarrow S$. Chunking creates a new production rule P', where a condition C maps to an action A.
- **ACT-R (Adaptive Control of Thought—Rational):** ACT-R distinguishes between declarative memory (facts) and procedural memory (productions). It uses utility theory to select productions and incorporates activation spreading for memory retrieval [Anderson, 2007].
    - **Formalism:** The activation A_i of a memory chunk i is given by $A_i = B_i + \sum_j W_{ji} S_{ji}$, where B_i is the base-level activation and the summation term represents associative activation. The utility U_p of a production p is calculated as $U_p = P_p G - C_p$, where P_p is the probability of success, G is the goal value, and C_p is the cost.

3.3. Connectionist and Neural-Symbolic Architectures

Deep learning models are powerful components for perception and pattern recognition but are weak in symbol manipulation and reasoning [LeCun et al., 2015]. Neural-Symbolic integration seeks to combine the strengths of both paradigms, using symbolic logic to guide neural networks and neural networks to ground symbolic representations [Garcez & Lamb, 2020].

4. Learning Paradigms and Knowledge Representation

4.1. Key Learning Paradigms

- **Reinforcement Learning (RL):** RL provides a general framework for goal-directed behaviour through Markov Decision Processes (MDPs) [Sutton & Barto, 2018].
    - **Formalism:** The Bellman equation defines the optimal value function: $V(s) = max_a \sum_{s',r} P(s',r \vee s,a)\left[r + \gamma V(s')\right]$.
- **Unsupervised and Self-Supervised Learning:** These paradigms leverage vast amounts of unlabeled data. Generative Models like GANs [Goodfellow et al., 2014], VAEs [Kingma & Welling, 2013], and Diffusion Models learn underlying data distributions.
    - **Formalism:** Variational Autoencoders (VAEs) maximize the Evidence Lower Bound (ELBO): $L(\theta,\phi;x) = E_{q_\phi(z \vee x)}\left[log p_\theta(x \vee z)\right] - D_{KL}(q_\phi(z \vee x) \vee p(z))$.
- **Meta-Learning and Lifelong Learning:** Meta-learning, or "learning to learn," aims to enable rapid adaptation [Finn et al., 2017]. Continual learning addresses "catastrophic forgetting" [Kirkpatrick et al., 2017].
    - **Formalism:** Elastic Weight Consolidation (EWC) mitigates forgetting by adding a penalty term to the loss function: $L(\theta) = L_B(\theta) + \sum_i \frac{\lambda}{2} F_i(\theta_i - \theta_{B,i})^2$, where F is the Fisher information matrix.

4.2. Knowledge Representation and Reasoning

- **Formal Logics and Ontologies:** First-Order Logic (FOL) and Description Logics (DLs) provide frameworks for structured knowledge [Baader et al., 2003].
- **Probabilistic Graphical Models (PGMs):** Bayesian Networks and Markov Random Fields represent uncertain knowledge and perform causal inference [Koller & Friedman, 2009].
    - **Formalism:** A Bayesian Network represents a joint probability distribution as: $P(X_1,\ldots,X_n) = \prod_{i=1}^{n} P(X_i \vee \mathrm{Pa}(X_i))$.
- **Knowledge Graphs and Causal Reasoning:** Knowledge Graphs represent facts as nodes and edges [Bordes et al., 2013]. Causal reasoning, using

frameworks like Judea Pearl's do-calculus, allows models to understand "why" an event occurred, moving beyond mere correlation [Pearl, 2009].

5. Hardware Enablement and Optimization Strategies

The theoretical frameworks for AGI are computationally voracious. Their practical implementation depends critically on powerful hardware and sophisticated optimization techniques.

5.1. Hardware Acceleration and Infrastructure

- **Specialized Processors:** The foundation of modern AI training lies in specialized hardware. Graphics Processing Units (GPUs) are the workhorses, with architectures designed for the massive parallel processing required by neural networks [NVIDIA, n.d. (1)]. Tensor Processing Units (TPUs) are custom-designed ASICs that offer even greater efficiency for large-scale training workloads [Google Cloud, n.d.].

- **Distributed Training:** For models that are too large or take too long to train on a single device, distributed training is essential. Data parallelism replicates a model across multiple devices, with each processing a different subset of data. Model parallelism splits a single large model across multiple devices. Hybrid approaches are often used for maximum efficiency [Jiegroup GenAI, n.d.].

- **Heterogeneous Computing Platforms:** Platforms like the NVIDIA Jetson AGX line integrate a multi-core CPU, a powerful GPU, and dedicated deep learning accelerators. This architecture is exceptionally well-suited for hybrid AGI models, allowing symbolic reasoning to run on the CPU while neural networks are accelerated by the GPU.

5.2. Algorithmic and Software Optimizations

- **Memory and Precision Optimization:** Mixed-precision training uses lower-precision formats like FP16, significantly reducing memory footprint and speeding up computation on compatible hardware [Neptune.ai, n.d.; Bharataameriya, 2025]. Gradient checkpointing trades computation for memory by recomputing activations during the backward pass instead of storing them, enabling the training of deeper networks [Bharataameriya, 2025].

- **Efficient Architectures:** The quadratic complexity of the self-attention mechanism in Transformers [Vaswani et al., 2017] is a major bottleneck. More efficient attention mechanisms like FlashAttention have been developed to reduce this computational burden [Brilworks, 2024]. MobileNets are an example of an efficient architecture designed for mobile vision applications [Howard et al., 2017].
- **Model Compression:** While often used for deployment, compression techniques impact training strategy. Pruning removes redundant weights, quantization reduces the numerical precision of weights, and knowledge distillation trains a smaller "student" model to mimic a larger "teacher" model, all resulting in smaller, faster models [Yugank.Aman, 2025].

5.3. Data Efficiency and Training Procedures

- **Transfer Learning and Fine-Tuning:** Instead of training from scratch, starting with a pre-trained model and fine-tuning it on a smaller, task-specific dataset saves immense computational resources [Brilworks, 2024; Nitor Infotech, n.d.].
- **Parameter-Efficient Fine-tuning (PEFT):** Techniques like LoRA (Low-Rank Adaptation) freeze most of a pre-trained model's weights and inject small, trainable matrices. This drastically reduces the number of trainable parameters, making the adaptation of large models highly efficient [Nitor Infotech, n.d.].
- **Data Curation and Learning Schedules:** High-quality training data is crucial. Active learning can be used to select the most informative data points for labeling, reducing the overall data required [Number Analytics, 2025]. Additionally, carefully tuning the learning rate during training using schedules can lead to faster convergence [Goodfellow et al., 2016].

6. Integrative Approaches and Future Directions

The most promising paths to AGI lie in the integration of the concepts discussed.

6.1. Hybrid Architectures and World Models

A deep dive into neuro-symbolic AI reveals methods to integrate symbolic planning with reinforcement learning. The execution of such models is made feasible by the heterogeneous compute resources of platforms like the Jetson AGX. A promising direction involves agents that learn internal simulations (world models) of their environment, a concept unified by the free energy principle [Friston, 2010].

6.2. Self-Improvement and Developmental AI

Inspired by human development, developmental AI involves the gradual acquisition of skills, often driven by intrinsic motivation [Oudeyer et al., 2007]. The ultimate goal is an AGI capable of recursively improving its own intelligence, a concept that leads to the idea of an "intelligence explosion" [Chalmers, 2010].

7. Challenges, Limitations, and Open Problems

The development of AGI faces numerous significant hurdles. These include the profound Common Sense Problem and the brittleness of current AI outside its training data. Data efficiency remains a major issue, as deep learning models require massive datasets compared to humans. The "black-box" nature of complex models creates a need for Interpretability and Explainability (XAI) [Gunning & Aha, 2019]. Furthermore, the immense computational resources required raise scalability and energy consumption concerns. Finally, the Ethical, Safety, and Societal Implications—most notably the alignment and control problems—are paramount [Bostrom, 2014].

8. Conclusion

The path to AGI is not through a single algorithm but through the integration of diverse architectural and mathematical frameworks, powered by increasingly capable and efficiently utilized hardware. This paper has outlined the key components, from cognitive architectures and learning paradigms to knowledge representation and the critical role of hardware and optimization. Future research must focus on sophisticated neuro-symbolic integration, advanced meta-learning, and robust causal inference. Crucially, the convergence of these software frameworks with advanced hardware platforms is creating a fertile ground for practical experimentation. This synergy allows for the deployment and testing of complex AGI concepts in real-world, embodied systems—a critical step towards realizing the promise of machines with human-like intelligence.

References

- Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe?* Oxford University Press.
- Baader, F., Calvanese, D., McGuinness, D. L., Nardi, D., & Patel-Schneider, P. F. (Eds.). (2003). *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press.
- Bharataameriya. (2025, February 17). Memory Optimization Techniques for Large Deep Learning Models. *Medium*. Retrieved from https://medium.com/@bharataameriya/memory-optimization-techniques-for-large-deep-learning-models-abf149588a74
- Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., & Yakhnenko, O. (2013). Translating Embeddings for Modeling Multi-relational Data. In *Advances in Neural Information Processing Systems 26* (pp. 2787–2795).
- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.
- Bringsjord, S., Bello, P., & Ferrucci, D. (2001). Creativity, the Turing Test, and the (Better) Lovelace Test. *Minds and Machines*, *11*(1), 3–27.
- Brilworks. (2024, September 9). 10 Ways to Improve AI Model Efficiency. Retrieved from https://www.brilworks.com/blog/improve-ai-model-efficiency/
- Chalmers, D. J. (2010). The Singularity: A Philosophical Analysis. *Journal of Consciousness Studies*, *17*(9-10), 7–65.
- Finn, C., Abbeel, P., & Levine, S. (2017). Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In *Proceedings of the 34th International Conference on Machine Learning* (PMLR 70, pp. 1126–1135).
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138.
- Garcez, A. d'Avila, & Lamb, L. C. (2020). Neurosymbolic AI: The 3rd Wave. *arXiv preprint arXiv:2012.05876*.
- Goertzel, B. (2012). The Robot College Student Test. In *AGI-12 Workshop: AGI-as-a-Science*.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative Adversarial Networks. In *Advances in Neural Information Processing Systems 27* (pp. 2672–2680).
- Google Cloud. (n.d.). TPU. Retrieved from https://cloud.google.com/tpu

- Gunning, D., & Aha, D. (2019). DARPA's Explainable Artificial Intelligence (XAI) Program. *AI Magazine*, *40*(2), 44–58.
- Harnad, S. (1990). The Symbol Grounding Problem. *Physica D: Nonlinear Phenomena*, *42*(1–3), 335–346.
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv preprint arXiv:1704.04861*.
- Jiegroup GenAI. (n.d.). 6. Efficient Training of Large Models. Retrieved from https://jiegroup-genai.readthedocs-hosted.com/en/latest/resource/
- Kho, Raymund K.D. et al., (2025). A Framework for Autonomous Marine Microplastic Detection: Integrating SWIR-Equipped Drones and Intelligent Buoys. Spectral Sensing. The Center for Research and Development Hong Kong (CRD-HK).
- Kingma, D. P., & Welling, M. (2013). Auto-Encoding Variational Bayes. *arXiv preprint arXiv:1312.6114*.
- Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., ... & Hadsell, R. (2017). Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, *114*(13), 3521–3526.
- Koller, D., & Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. MIT Press.
- Laird, J. E. (2012). *The Soar Cognitive Architecture*. MIT Press.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.
- Legg, S., & Hutter, M. (2007). A Collection of Definitions of Intelligence. In *Advances in Artificial General Intelligence* (pp. 17–24). IOS Press.
- McCarthy, J., & Hayes, P. J. (1969). Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence*, *4*, 463-502.
- McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*.
- Neptune.ai. (n.d.). Deep Learning Model Optimization Methods. Retrieved from https://neptune.ai/blog/deep-learning-model-optimization-methods
- Newell, A. (1990). *Unified Theories of Cognition*. Harvard University Press.
- Nitor Infotech. (n.d.). Training Large Language Models (LLMs): Techniques and Best Practices. Retrieved from https://www.nitorin-

fotech.com/blog/training-large-language-models-llms-techniques-and-best-practices/

- NVIDIA. (n.d.). NVIDIA Tensor Cores. Retrieved from relevant NVIDIA documentation.
- Number Analytics. (2025, June 11). Maximizing AI Efficiency with Active Learning. Retrieved from https://www.numberanalytics.com/blog/maximizing-ai-efficiency-active-learning
- Oudeyer, P.-Y., Kaplan, F., & Hafner, V. V. (2007). Intrinsic Motivation Systems for Autonomous Mental Development. *IEEE Transactions on Evolutionary Computation*, *11*(2), 265–286.
- Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359.
- Pearl, J. (2009). *Causality: Models, Reasoning, and Inference* (2nd ed.). Cambridge University Press.
- Pfeifer, R., & Bongard, J. C. (2006). *How the Body Shapes the Way We Think: A New View of Intelligence*. MIT Press.
- Searle, J. R. (1980). Minds, Brains, and Programs. *Behavioral and Brain Sciences*, *3*(3), 417–424.
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (2nd ed.). MIT Press.
- Thrun, S., & Pratt, L. (Eds.). (2012). *Learning to Learn*. Springer Science & Business Media.
- Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, *59*(236), 433–460.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems, 30*.
- Wozniak, S. (2010). In *The University of Michigan News Service interview, "Wozniak: 'The computer should be the servant'"*.
- Yugank .Aman. (2025, February 14). LLM Model Optimization Techniques and Frameworks. *Medium*. Retrieved from https://medium.com/@yugank.aman/llm-model-optimization-techniques-and-frameworks-e21d57744ca1