

PhD Research Proposal: Advancing General Artificial Intelligence (AGI) Through Integrative Cognitive Architectures

Applicant: Karthikeyan Kannan

1. Introduction

Artificial General Intelligence (AGI) represents the future of AI, aiming to develop systems capable of performing any intellectual task that a human can do. Unlike narrow AI, which excels in specific tasks, AGI promises to exhibit human-like adaptability, reasoning, and problem-solving across a broad range of domains. However, achieving AGI is still an open challenge.

This proposal outlines a research plan to create **Integrative Cognitive Architectures (ICA)** that incorporate elements from symbolic AI, neural networks, and neuroscience to build flexible and autonomous AGI systems. The research will explore the following key areas:

- **Cross-domain learning and generalization**, enabling AGI to adapt knowledge and skills from one domain to another.
 - **Uncertainty management**, ensuring AGI systems can function effectively in dynamic and unpredictable environments.
 - **Neuroscience-inspired mechanisms**, providing insights into human-like cognition, memory, and learning.
-

2. Research Questions

The primary research questions to be addressed in this study are:

1. **How can an Integrative Cognitive Architecture be designed to mimic human cognition and perform across multiple domains?**
 2. **What strategies can enhance AGI's ability to generalize knowledge across various tasks and environments?**
 3. **How can AGI systems effectively manage uncertainty and make decisions under incomplete information?**
 4. **What neuroscience-based mechanisms can be incorporated to improve AGI's memory, reasoning, and adaptability?**
-

3. Research Objectives

The objectives of this research are:

1. **Develop a novel Integrative Cognitive Architecture (ICA)** that mimics human-like reasoning, decision-making, and learning across diverse tasks.
 2. **Enhance AGI's generalization capability** through transfer learning techniques that enable systems to adapt knowledge from one domain to another.
 3. **Integrate uncertainty management** techniques, such as probabilistic reasoning, to allow AGI systems to operate effectively in dynamic and incomplete environments.
 4. **Incorporate neuroscience principles**, such as memory consolidation and neural plasticity, to enhance AGI's long-term learning, memory, and problem-solving capabilities.
-

4. Literature Review

The following research streams are key to the foundation of this project:

1. **Cognitive Architectures:** Traditional architectures, such as SOAR and ACT-R, simulate human cognition but struggle with scalability and adaptability (Laird et al., 2017).
 2. **Neural Networks and Deep Learning:** While these models have achieved success in narrow AI, their limited generalization across diverse tasks is a significant barrier to AGI (LeCun et al., 2015).
 3. **Hybrid Systems:** Combining symbolic AI with neural networks offers the potential for systems that balance flexibility and interpretability (Marcus, 2020).
 4. **Neuroscience-Inspired AI:** Cognitive neuroscience has provided valuable insights into learning, memory, and decision-making, offering models that could enhance AGI (Hassabis et al., 2017).
-

5. Methodology

This research will span three years, divided into the following phases:

Year 1: Development of Integrative Cognitive Architecture (ICA)

- **Framework Design:** Develop the ICA framework that integrates symbolic reasoning, neural networks, and neuroscience principles.
- **Basic Module Implementation:** Design initial modules for perception, memory, and reasoning.
- **Simulation and Testing:** Validate the framework on basic tasks (e.g., pattern recognition, logical reasoning).

Year 2: Generalization and Transfer Learning

- **Implement Generalization Mechanisms:** Introduce methods for multi-modal learning and generalization across domains.
- **Develop Transfer Learning Strategies:** Enable AGI systems to apply knowledge from one domain (e.g., language) to another (e.g., robotics).
- **Experimental Evaluation:** Test the system's generalization abilities and compare with baseline models.

Year 3: Uncertainty Management and Neuroscience-Inspired Enhancements

- **Incorporate Uncertainty Management:** Integrate probabilistic reasoning to handle incomplete data and decision-making under uncertainty.
- **Neuroscience-Based Enhancements:** Introduce memory consolidation and neural plasticity models to enhance cognitive flexibility and long-term memory retention.
- **Final Testing:** Evaluate the final ICA on complex tasks in multiple domains (e.g., robotics, natural language processing, decision-making).

Thesis Writing and Defense

- Prepare and submit the dissertation, documenting the research process, results, and conclusions.

6. Expected Contributions

This research will make the following contributions:

1. **Development of an Integrative Cognitive Architecture (ICA)** that combines symbolic reasoning, neural networks, and neuroscience-based models for AGI.
 2. **Advancement of AGI generalization techniques**, allowing systems to transfer knowledge across tasks without needing retraining.
 3. **Enhanced uncertainty management capabilities**, enabling AGI to make decisions in dynamic environments with incomplete information.
 4. **Incorporation of neuroscience-inspired techniques** to improve AGI's cognitive abilities, particularly in memory, learning, and adaptability.
-

7. Timeline

Research Activity	Year 1	Year 2	Year 3
Design and Development of ICA Framework	✓		
Implementation of Generalization Mechanisms		✓	
Transfer Learning and Task Adaptation		✓	
Uncertainty Management Integration			✓
Neuroscience-Inspired Enhancements			✓
Simulations and Testing	✓	✓	✓
Thesis Writing and Defense			✓

8. Resources Required

- Access to **high-performance computing resources** for training and testing machine learning models.
 - Collaboration with **cognitive science and neuroscience departments** for interdisciplinary insights and expertise.
 - Datasets and simulation environments (e.g., robotics tasks, natural language processing datasets) to evaluate AGI's generalization and learning capabilities.
-

9. References

1. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
2. Marcus, G. (2020). The next decade in AI: Four steps towards robust artificial intelligence. *arXiv preprint arXiv:2002.06177*.
3. Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron*, 95(2), 245-258.
4. Laird, J. E., Lebiere, C., & Rosenbloom, P. S. (2017). A standard model of the mind: Toward a common computational framework across artificial intelligence, cognitive science, neuroscience, and robotics. *AI Magazine*, 38(4), 13-26.
5. Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433-460.
6. Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85-117.
7. Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40.
8. Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.
9. Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus, and Giroux.
10. Pearl, J. (2019). *The Book of Why: The New Science of Cause and Effect*. Basic Books.
11. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
12. Clark, A. (2015). Surfing uncertainty: Prediction, action, and the embodied mind. *Oxford University Press*.

13. Mnih, V., Kavukcuoglu, K., Silver, D., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529-533.
14. Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in Cognitive Sciences*, 11(7), 299-306.
15. Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, 350(6266), 1332-1338.