

Research Proposal for Caltech SURF Program by Malhar Ubhe

Title: Towards Artificial General Intelligence - A Multi-Agent Framework Approach

Artificial General Intelligence (AGI) is one of the most crucial stages in AI research, wherein we aim to create systems capable of human-like cognitive abilities, which includes skills like reasoning, critical thinking, adaptability, learning and a problem-solving approach towards diverse tasks. At present studies have achieved success in developing Artificial Narrow Intelligence (ANI) systems, which are mainly used to carry out domain specific tasks. My research interests lie in this transition from Narrow Intelligence to Generalized Intelligence wherein, I believe that AGI must demonstrate versatility and autonomy. Despite remarkable progress in deep learning and reinforcement learning, existing AI models remain fundamentally constrained by their task-specific design and lack of generalization.

With this view, I believe that AGI can be achieved through a multi-agent framework, where specialized domain specific AI agents collaborate in a dynamic environment. This approach impersonates human intelligence, where multiple cognitive subsystems interact to solve complex problems. By this approach, we can explore AI capabilities, cooperation between agents and develop a scalable learning mechanism that ANI systems struggle to achieve. Another major challenge that we will have to address before is which tasks can be termed as AGI Tasks. This task definition would help us to evaluate the effectiveness of our proposed systems which could lead us towards AGI.

Research Objectives:- This study aims to investigate the feasibility of a multi-agent framework as a pathway toward AGI development. Specifically, the research will:

- Define a set of AGI-relevant tasks and assess the performance of current AI models on these benchmarks.
- Design an architecture in which multiple ANI agents interact and collaborate to achieve General Intelligence.
- Develop protocols for inter-agent communication, coordination, and knowledge-sharing.
- Evaluate the novel problem-solving capabilities of the proposed system on benchmark tasks.
- Identify fundamental architectural and computational constraints that must be addressed for AGI to become viable.

Approach:- To achieve the research objectives, the following approach will be adopted:

- Defining AGI Tasks: Selection of a diverse set of benchmark tasks that test reasoning, critical thinking, adaptability, learning and collaboration.
- Multi-Agent Infrastructure: Using available tools like Autogen, Langchain, Phidata, etc or developing new multi-agent infrastructure workflows and integrating them with LLMs and creating a playground of agents to solve various AGI Tasks. Also establish communication protocols for agents to share knowledge and coordinate actions.
- Developing Learning strategies: Applying reinforcement learning techniques to enhance adaptability and enable agents to dynamically change their strategies depending upon the past interactions or scenarios.
- Implementation and Simulation: Developing and testing the framework within an AI simulation environment, where agents will be trained to solve various tasks.
- Evaluation and Analysis: Benchmarking the multi-agent framework against existing AI models, assessing adaptability to unseen challenges and analyzing emergent cooperative behaviors and learning dynamics.
- Overcoming Challenges: Addressing scalability issues through efficient methodologies and robustness testing and adversarial training.

Work Plan:- A structured timeline will be followed to ensure systematic progress:

- Week 1-2: Literature review, task identification, and initial framework design.
- Week 3-4: Implementation of individual agents and inter-agent communication mechanisms.
- Week 5-6: System integration, preliminary simulations, and debugging.
- Week 7-8: Performance evaluation, identification of key constraints, and iterative refinement.
- Week 9: Report documentation, final oral presentation and research paper preparation.

References:-

- Generative AI vs. AGI: Cognitive Strengths and Weaknesses of Modern LLMs - By Ben Goertzel
- A Foundational Architecture for Artificial General Intelligence - By Stan Franklin
- Artificial General Intelligence: Concept, State of the Art, and Future Prospects - By Ben Goertzel
- Sparks of Artificial General Intelligence: Early experiments with GPT-4 (Microsoft Research)
- The AGI Containment Problem - By James Babcock, Janos Kramar, and Roman Yampolskiy
- Towards artificial general intelligence via a multimodal foundation model.
- When brain-inspired AI meets AGI.
- Contemporary Approaches to Artificial General Intelligence - By Cassio Pennachin and Ben Goertzel.