Indian Journal of Computer Science and Technology
https://www.doi.org/10.59256/indjcst.20250402043
Volume 4, Issue2 (May-August 2025), PP: 279-289.
www.indjcst.com
ISSN No: 2583-5300

Agentic Approach in the Quest for AGI

Mahesh Basavaraj
Assistant Professor, Department of Computer Science – Data Science, Dayananda Sagar Academy of Technology and Management, Bangalore, Karnataka, India.

***To Cite this Article**: Mahesh Basavaraj, "Agentic Approach in the Quest for AGI", Indian Journal of Computer Science and Technology, Volume 04, Issue 02 (May-August 2025), PP: 279-289.*

Abstract: *Artificial General Intelligence (AGI) aims to create machine intelligence rivaling human cognitive abilities across all domains. Current approaches struggle with common-sense reasoning, cross-domain knowledge transfer, and consistent multi-step task performance. This paper investigates a collaborative framework using AI agents built on Large Language Models (LLMs) to achieve AGI. By harnessing LLMs' strengths in language processing, reasoning, and knowledge synthesis, we propose a multi-agent system where specialized agents work together to exhibit emergent general intelligence. We analyze LLM technologies, multi-agent coordination, and existing implementations like Google's SayCan, OpenAI's multi-agent systems, ChatDev, and Devin. Our findings suggest that collaborative LLM-based agent systems are the most promising path to AGI, despite challenges in embodied cognition, safety alignment, and true understanding versus pattern matching. We outline a roadmap for future research and key open questions in developing AGI through collaborative architectures.*

Key Words: *Artificial General Intelligence, Large Language Models, AI Agents, Multi-Agent Systems, Collaborative Intelligence, Machine Learning.*

I.INTRODUCTION

Artificial General Intelligence (AGI) refers to machine intelligence capable of human-level or superhuman performance across diverse domains. Unlike narrow AI, AGI can adapt, transfer knowledge, and solve novel problems flexibly.

Historically, symbolic AI struggled with real-world ambiguity, while neural network-based models required vast data and lacked generalization. Recent breakthroughs in Large Language Models (LLMs) like GPT-4, Claude, and Gemini have significantly advanced language understanding, reasoning, and code generation. However, LLMs still lack autonomy, persistent learning, and real-world interaction.

This paper proposes a collaborative multi-agent framework built on LLMs as the most viable path to AGI. By integrating specialized agents—each handling functions like reasoning, memory, planning, execution, and communication—we combine LLM strengths with agentic autonomy and interaction capabilities. This architecture enables error correction, goal-driven behavior, and emergent intelligence, overcoming limitations of standalone systems and advancing toward practical AGI.

To contextualize this proposal, it is essential to first understand the evolution of AGI research and the specific capabilities of modern LLMs that make such a framework both necessary and viable.

II.CURRENT STATE OF AGI AND LLM RESEARCH

2.1 Evolution of AGI Research Approaches

The quest for AGI has evolved through several key phases, each offering valuable progress while exposing critical limitations. **Symbolic AI**, dominant in the early stages, relied on rule-based systems and explicit knowledge representation. While effective in structured environments such as expert systems, these approaches were brittle and failed to scale to real-world complexity.

Connectionist approaches, particularly deep learning, have since driven major advances in pattern recognition, language processing, and decision-making. These models excel at learning from large datasets and have achieved remarkable results in specific tasks. However, they often lack the flexibility, generalization, and transfer learning abilities essential for true general intelligence.

Hybrid or neurosymbolic approaches attempt to combine the structured reasoning of symbolic AI with the adaptive learning of neural networks. While promising in theory, these systems face significant challenges in integration, making it difficult to achieve the seamless coordination and robustness required for AGI-level performance.

Building on these historical foundations, the recent emergence of Large Language Models marks the next significant paradigm shift in this quest.

2.2 Large Language Model Capabilities and Limitations

Contemporary LLMs represent a significant leap forward in AI capabilities, demonstrating sophisticated performance across multiple cognitive domains:

Language Understanding and Generation : Modern LLMs excel in contextual understanding, semantic reasoning, and generating coherent text across topics, enabling sophisticated dialogue and response generation.
Reasoning and Problem-Solving: LLMs demonstrate strong logical and mathematical reasoning, with techniques like chain-of-thought prompting enabling step-by-step problem-solving.
Knowledge Integration and Synthesis: These models can connect concepts across domains, synthesizing information to generate novel insights—key for general intelligence.
Code Generation and Technical Skills: LLMs are proficient in programming, algorithm design, and technical tasks, indicating capabilities beyond language processing.
Multi-Modal Integration: Models like GPT-4V and Gemini integrate visual and textual understanding, expanding their grounded comprehension abilities.

Limitations of Current LLMs
Lack of Persistent Learning: LLMs cannot learn or update their knowledge post-deployment, limiting adaptability and continuous improvement.
Inconsistency and Hallucination: They sometimes produce inaccurate yet plausible content, posing reliability challenges for critical tasks.
Limited Planning Horizon: LLMs struggle with sustained long-term planning and managing complex goals over time.
Absence of Embodied Experience: Without sensory input or physical interaction, LLMs lack real-world grounding in causality and dynamics.
Context Window Limitations: Despite improvements, they still face challenges handling very long sequences, affecting coherence over extended inputs.

To overcome these inherent limitations, a new architectural approach is required. By structuring LLMs within an agentic framework, we can introduce the necessary components for persistent learning, planning, and real-world interaction

III.AI AGENTS: FOUNDATIONS FOR COLLABORATIVE INTELLIGENCE

3.1 Theoretical Framework for LLM-Based Agents

Here's the **condensed version** of the conceptual architecture of LLM-based agents, structured with each module's key role:

Conceptual Architecture of LLM-Based Agents
Brain Module: Core LLM component handling language interaction, knowledge integration, memory management, reasoning, and high-level decision-making.
Perception Module: Processes multimodal inputs—text, images, audio, and structured data—using vision encoders, audio systems, and sensor interfaces for environmental awareness.
Action Module: Enables digital and physical actions such as tool usage, API calls, file manipulation, robot control, and communication with users or agents.
Memory Module: Maintains context across sessions, storing and retrieving information to support long-term interaction and overcome base LLM limitations.
Learning Module: Supports continuous improvement through experience analysis, feedback integration, and adaptive capability development.
This summary keeps the key components and their functions clear and compact while retaining the original intent.

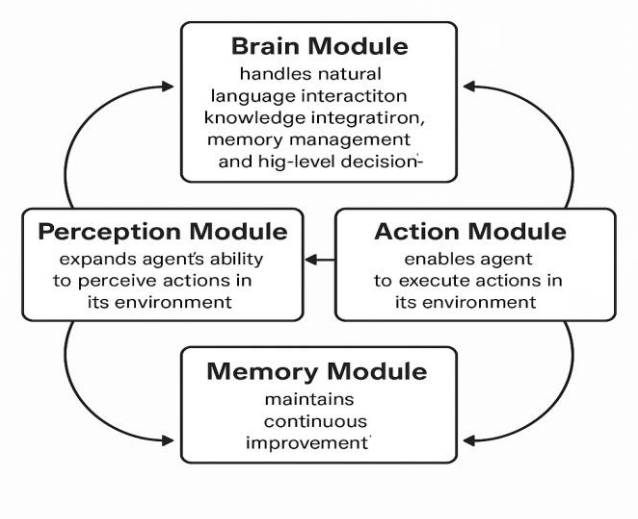

Fig-1: Theoretical Framework for LLM-Based Agents.

3.2 Why LLMs are Ideal Agent Foundations

LLMs exhibit several key properties that make them exceptionally well-suited as the foundation for intelligent agents:
Autonomy: LLMs demonstrate the ability to operate independently, generate creative solutions, and exhibit goal-directed

behavior without constant human oversight. They can interpret instructions, make decisions, and pursue objectives with minimal intervention.

Reactivity: These models can rapidly respond to environmental changes, processing new information and adjusting their behavior accordingly. Their ability to understand context and adapt their responses makes them highly reactive to changing circumstances.

Pro-activeness: LLMs exhibit goal-oriented behavior through reasoning and planning capabilities. They can break down complex objectives into actionable steps, anticipate future needs, and take initiative in pursuing their goals.

Social Ability: The natural language interface of LLMs enables sophisticated interaction with other agents and humans. They can engage in complex dialogue, negotiate, collaborate, and communicate effectively across diverse contexts.

Knowledge Integration: LLMs excel at combining information from multiple sources, synthesizing knowledge across domains, and applying learned concepts to new situations. This capability is essential for agents operating in complex, multi-faceted environments.

When these inherent properties are harnessed within a structured agentic framework, they give rise to a new set of powerful, emergent capabilities.

3.3 Enhanced Capabilities Through Agentic Architectures

The transformation of LLMs into autonomous agents unlocks several enhanced capabilities crucial for AGI development:

Tool Usage and External Interaction: LLM-based agents can learn to use external tools, APIs, and services to accomplish complex tasks. This capability extends their influence beyond text generation to real-world problem-solving and task execution.

Multi-Modal Understanding: Integration of perception modules allows agents to process and understand information from multiple sensory modalities, providing a more comprehensive understanding of their environment.

Persistent Context and Memory: Advanced memory systems enable agents to maintain long-term context, learn from past experiences, and build upon previous interactions to improve future performance.

Planning and Strategy: Sophisticated planning modules allow agents to develop and execute complex strategies, manage multiple objectives, and adapt their approaches based on changing circumstances.

Collaborative Coordination: Agents can communicate, collaborate, and coordinate with other agents to achieve shared objectives, enabling collective intelligence that exceeds individual capabilities.

While a single, highly capable agent represents a significant step forward, the true potential for achieving AGI lies in moving from individual intelligence to collective, collaborative intelligence. This requires exploring architectures where multiple agents can work together.

IV.MULTI-AGENT SYSTEMS FOR COLLABORATIVE AGI DEVELOPMENT

4.1 Cooperative Multi-Agent System Architectures

Cooperative Multi-Agent Systems (MAS) represent a powerful framework for achieving AGI through collaborative intelligence. In these systems, multiple AI agents work together toward shared objectives, combining their specialized capabilities to solve complex problems that would be challenging for individual agents.

Distributed Specialization: Different agents can specialize in specific cognitive functions or domain expertise while contributing to collective problem-solving. For example, one agent might specialize in logical reasoning, another in creative generation, and a third in fact verification and validation.

Consensus-Based Decision Making: Multiple agents can contribute to decision-making processes, with disagreements resolved through evidence-based discussion, voting mechanisms, or hierarchical arbitration. This approach improves reliability and reduces the impact of individual agent errors.

Knowledge Sharing and Integration: Agents can share learned experiences, insights, and knowledge, collectively building a more comprehensive understanding of their operational domain. This collaborative learning accelerates progress and reduces redundant effort.

Error Correction and Validation: Multiple agents can cross-check each other's work, identifying errors, inconsistencies, and potential improvements. This multi-agent validation significantly improves system reliability and accuracy.

4.2 Hierarchical Multi-Agent Systems

Hierarchical architectures introduce additional structure and coordination mechanisms to multi-agent systems:

High-Level Coordination: Meta-agents or coordinator agents manage overall system behavior, allocating tasks, resolving conflicts, and ensuring coherent system-wide objectives are maintained.

Specialized Task Decomposition: Complex objectives are broken down into manageable subtasks, with appropriate agents assigned to each component based on their specialized capabilities.

Adaptive Resource Allocation: The system can dynamically allocate computational resources and agent attention based on current priorities and requirements.

Emergent Intelligence: Sophisticated behaviors and capabilities emerge from the hierarchical interactions between agents, potentially leading to intelligence that exceeds the sum of individual components.

While these hierarchical and cooperative architectures provide a theoretical blueprint for collaborative AGI, their practical implementation presents a unique set of challenges that must be addressed.

4.3 Implementation Challenges and Solutions

Distributed Decision-Making: Coordinating decisions across multiple autonomous agents requires sophisticated protocols for

information sharing, consensus building, and conflict resolution. Solutions include voting mechanisms, reputation systems, and hierarchical decision structures.

Communication Overhead: Extensive inter-agent communication can create bottlenecks and reduce system efficiency. Optimizations include selective information sharing, compression techniques, and asynchronous communication protocols.

Safety and Explainability: Multi-agent systems must maintain safety constraints while providing transparency into their decision-making processes. This requires robust monitoring systems, explainable AI techniques, and fail-safe mechanisms.

Scalability: As systems grow in complexity and number of agents, maintaining performance and coordination becomes increasingly challenging. Solutions include modular architectures, distributed processing, and adaptive scaling mechanisms.

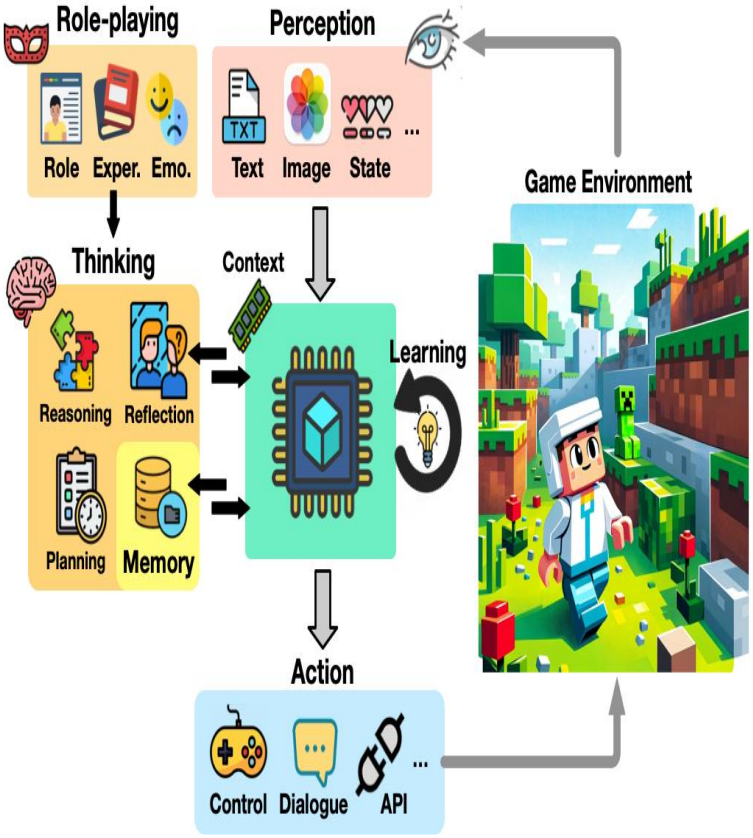

Fig-2: The conceptual architecture of Agentic ecosystem.

At each step, the perception module perceives the multimodal information from the game environment, including textual, images, symbolic states, and so on. The agent retrieves essential memories from the memory module and take them along with perceived information as input for thinking (reasoning, planning, and reflection), enabling itself to formulate strategies and make informed decisions. The role-playing module affects the decision-making process to ensure that the agent's behavior aligns with its designated character. Then the action module translates generated action descriptions into executable and admissible actions for altering game states at the next step. Finally, the learning module serves to continuously improve the agent's cognitive and inference abilities through the accumulated experience

Synergistic Potential: AI Agents built on LLMs for AGI Development:

a conceptual framework for constructing LLM-based agents comprising three key components: brain, perception, and action. The brain module, built primarily with a large language model, handles core functions like natural language interaction, knowledge integration, memory management, reasoning, and decision-making. The perception module expands the agent's ability to perceive and comprehend multimodal information from the environment, including text, audio, and visuals. The action module enables the agent to execute embodied actions, use tools, and influence its surroundings.

Agent Functionalities (Left Side):

Perception: This refers to the agent's ability to gather information about its environment. This could involve sensors for physical agents (robots) or processing information from text and code for LLMs.

Action: Based on its perception, the agent can take actions in the environment. Physical agents might move or manipulate objects, while LLMs might generate text, translate languages, or write different creative content.

Learning: Through interacting with the environment and receiving feedback, the agent learns and improves its ability to perform tasks.

Adaptation: The agent can adapt its behavior based on new information or experiences.

These functionalities work together in a cycle. The agent perceives, takes actions, learns from the results, and adapts its future behavior.

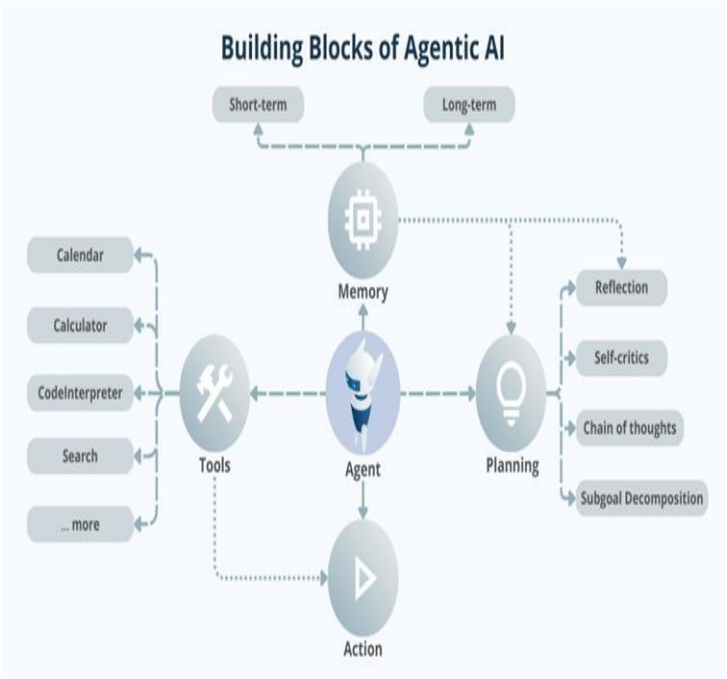

Fig-3: The building blocks of Agentic System.

Components of the LLM-based Agent System (Right Side):
Large Language Models (LLMs): These are AI models trained on massive amounts of text data. They excel in tasks like reasoning, communication, and knowledge representation, which are crucial for an AGI system. Knowledge Base: This component stores information that the agents can access and utilize. It can include factual knowledge, past experiences, and learned strategies.Dialogue Component: This allows communication between AI agents within the system and potentially with humans. This is important for information sharing, collaboration, and task coordination.
API (Application Programming Interface): This enables interaction between the multi-agent system and the external world. The API allows the system to receive information from the real world and potentially take actions within it.The two sides are interconnected. The functionalities of the agents (perception, action, learning, adaptation) rely on the capabilities provided by the LLM components (reasoning, communication, knowledge representation). This combined approach aims to create a more robust and intelligent AI system capable of complex tasks.

Advantages of Agentic approach for AGI:

AI agents built with large language models (LLMs) are becoming more intelligent and versatile. These agents can now tackle new tasks by following clear instructions, without needing specific training for each one. They can also learn from just a few examples, thanks to in-context learning.

Another big improvement is how these agents perceive information. They can now go beyond text and understand visual and auditory input. For sights, image encoders turn images into a format LLMs can comprehend. Sounds are processed by existing audio models and then integrated with the LLMs. This lets agents grasp information from the real world through multiple senses.

AI agents, with their core functionalities of perception, action, learning, and adaptation, offer a versatile framework for developing AGI. Here, we explore how AI agents built upon LLMs can contribute to this endeavour.
Enhanced Communication and Knowledge Sharing: LLMs excel at processing and understanding language. This allows the AI agents built upon them to facilitate communication and knowledge sharing between diverse agents within a multi-agent system. By sharing information and learned experiences, agents can collectively build a more comprehensive understanding of the world and approach complex problems collaboratively.
Reasoning and Problem-Solving with Language Foundation: LLMs can provide a foundation for reasoning and problem-solving within AI agents. By processing information and generating potential solutions through language, the LLM component can assist the AI agent in navigating complex situations and making informed decisions.
Transfer Learning Across Tasks and Domains: LLMs demonstrate impressive transfer learning abilities. By leveraging this capability within AI agents, different agents specializing in specific tasks (e.g., vision, robotics, language) can collectively contribute to achieving a broader goal, accelerating progress towards AGI.

V.MULTI-AGENT SYSTEMS FOR COLLABORATIVE AGI DEVELOPMENT

This section remains largely the same as before, focusing on exploring different architectures for utilizing AI agents built on LLMs in AGI development, including Cooperative Multi-Agent Systems (MAS) and Hierarchical MAS. It will also discuss the challenges associated with implementing these approaches, such as distributed decision-making, communication overhead, and ensuring safety and explainability within the system.

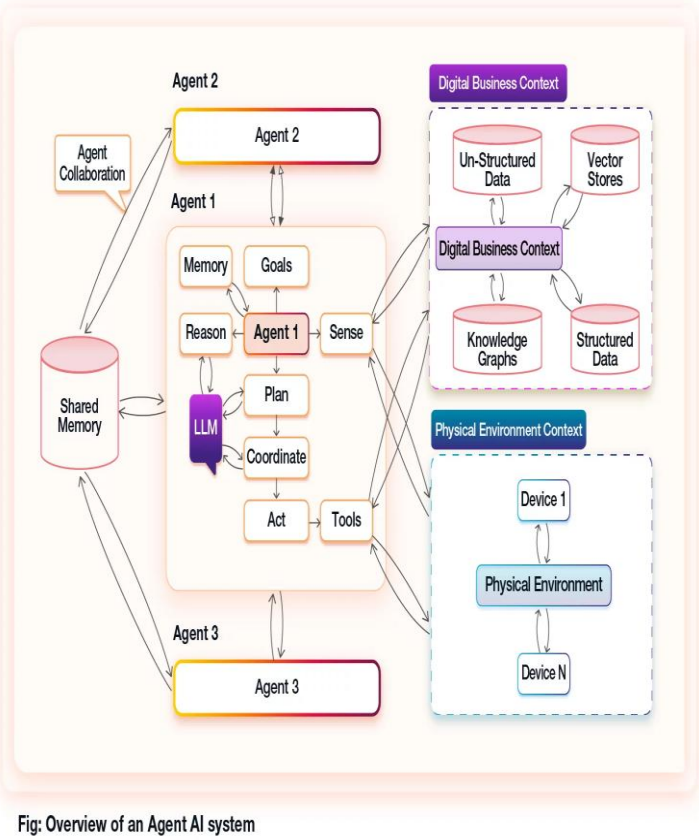

Fig-4: The conceptual architecture of Multi-Agent Agentic System

To better understand how these challenges are being tackled, it is useful to examine existing research and the current landscape of both open-source and commercial implementations.

VI.EXISTING RESEARCH AND CASE STUDIES

6.1 Open Source Agent Platforms

The open-source ecosystem continues to expand with robust platforms designed for LLM-based multi-agent collaboration:

- AutoGPT and LangChain remain pivotal in developing autonomous agent systems. These platforms demonstrate capabilities in task decomposition, memory-driven reasoning, and tool utilization.
- LangChain's emphasis on structured memory and environment control pairs well with AutoGPT's recursive task loop strategy, highlighting practical planning and execution for autonomous goals (Barua).
- CrewAI, AutoGen, and CAMEL present evolving open frameworks tailored for collaborative MAS deployment. These tools specialize in inter-agent communication, task negotiation, and coordinated decision-making, reinforcing scalable team-based autonomy (Tran et al.).

6.2 Commercial and Research Implementations

- ChatDev showcases a structured MAS workflow applied to software development. In this environment, role-specific agents (e.g., coder, reviewer, tester) emulate human team dynamics, demonstrating how distributed agents can collaborate to build and ship software systems. This model illustrates real-world applicability of agent specialization and collaborative execution (Åström & Winoy).
- Devin, the autonomous AI software engineer, exemplifies the integration of planning, reasoning, debugging, and system design in a single agent. It represents the frontier of autonomous technical cognition, where an agent can iteratively refine solutions, interact with APIs, and manage complete development cycles.
- Google's SayCan Project: This groundbreaking research demonstrates how LLMs can serve as high-level controllers for robotic systems. The project shows how natural language instructions can be translated into sequences of robot actions, bridging the gap between language understanding and physical world interaction. SayCan represents a significant step toward embodied AI agents that can understand and execute complex real-world tasks.
- Multi-Agent Collaboration Mechanisms explored in academic surveys emphasize the importance of structured communication channels, emergent behavior monitoring, and ethical alignment in complex MAS environments. These implementations directly address AGI-relevant goals of scalability, safety, and goal alignment (Tran et al.).

Fig 5: Google SayCan in action

6.3 Emerging Platforms and Applications

- AutoGen and CAMEL represent novel paradigms in collaborative agent design, focusing on role-enforced communication protocols and consensus-driven decision-making among LLM-based agents.
- Experimental MAS Testbeds such as those adapted from OpenAI's original Universe framework continue to serve as simulation arenas for testing cooperation, adversarial strategies, and system robustness under dynamic constraints.

These studies underline that while MAS holds transformative promise for AGI development, realizing this vision demands continued progress in coordination protocols, transparency mechanisms, and safe autonomy design.

6.4 Analysis of Current Implementations

Current implementations demonstrate several key insights about LLM-based agents for AGI development:

Capability Emergence: When LLMs are enhanced with agentic capabilities, they exhibit behaviors and problem-solving abilities that exceed their base model capabilities.

This suggests that the agentic framework unlocks latent potential in LLMs.

Collaboration Benefits: Multi-agent systems consistently demonstrate superior performance compared to single-agent approaches in complex tasks. The collaborative approach provides error correction, specialized expertise, and creative problem-solving that individual agents cannot achieve.

Tool Integration Success: LLM-based agents show remarkable ability to learn and use external tools, APIs, and systems. This capability significantly extends their problem-solving reach and demonstrates the potential for real-world interaction.

Limitations and Challenges: Current implementations also reveal significant challenges, including consistency issues, planning limitations, and difficulties with long-term objective management. These challenges highlight areas requiring further research and development.

This analysis of current platforms reveals both the promise and the shortcomings of existing approaches. Based on these insights, we can now propose a more comprehensive and structured framework designed to leverage these strengths while addressing the identified gaps

VII.FRAMEWORK FOR COLLABORATIVE AGI DEVELOPMENT

7.1 Proposed Architecture

Based on our analysis of current research and implementations, we propose a comprehensive framework for AGI development through collaborative LLM-based agents:

Foundation Layer: Advanced LLMs serve as the cognitive foundation for all agents, providing natural language understanding, reasoning capabilities, and knowledge synthesis. This layer includes models like GPT-4, Claude, Gemini, and their successors.

Specialized Agent Layer: A collection of specialized agents, each optimized for specific cognitive functions:

- **Planning Agent:** Responsible for high-level goal decomposition, strategic planning, and objective management
- **Reasoning Agent:** Specialized in logical analysis, mathematical computation, causal inference, and problem-solving
- **Memory Agent:** Manages long-term knowledge storage, retrieval, context maintenance, and learning integration
- **Execution Agent:** Handles interaction with external tools, APIs, systems, and real-world interfaces
- **Validation Agent:** Performs error checking, fact verification, consistency analysis, and quality assurance
- **Learning Agent:** Continuously improves system performance through experience analysis and capability enhancement
- **Communication Agent:** Manages inter-agent communication, coordination, and information sharing

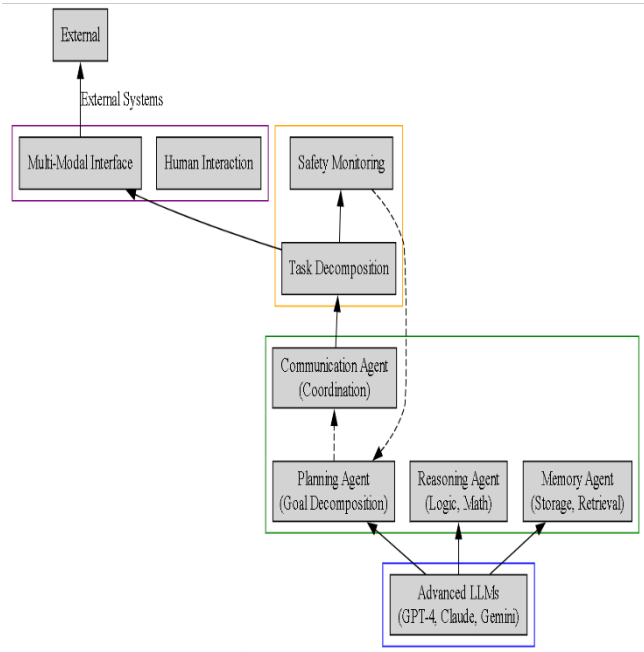

Fig 6. Proposed Architecture for LLM based AI agent system.

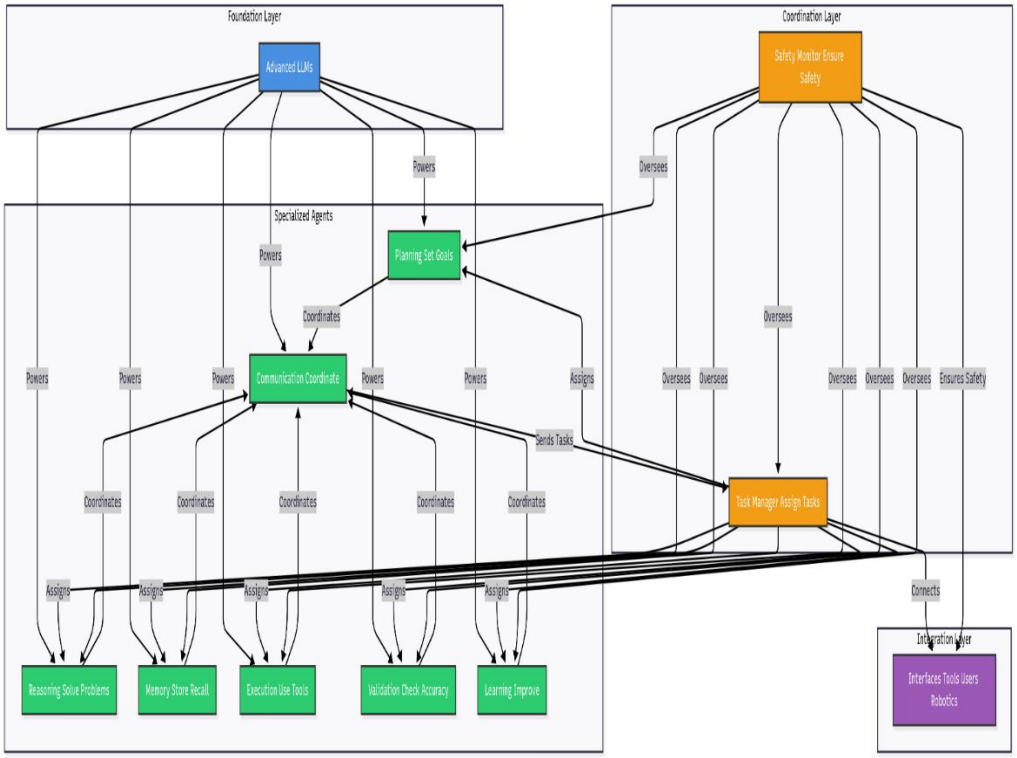

Fig 7: Layers of the Proposed Architecture for LLM based AI agent system and the interaction flowchart

Coordination Layer: Sophisticated mechanisms for managing agent interactions, task allocation, and system-wide behavior:

- **Task Decomposition Engine:** Breaks complex objectives into manageable subtasks and assigns them to appropriate agents
- **Consensus Formation System:** Resolves disagreements and builds agreement through evidence-based discussion and voting
- **Resource Management System:** Allocates computational resources, attention, and priority based on current needs
- **Safety Monitoring System:** Ensures all agent actions comply with safety constraints and human values

Integration Layer: Components that connect the agent system to the external world:

- **Multi-Modal Interface:** Processes visual, auditory, and sensory information from the environment
- **Tool Integration Framework:** Enables agents to use external software, APIs, and services
- **Human Interaction Interface:** Facilitates communication and collaboration with human users
- **Real-World Interface:** Connects to physical systems, robotics, and embodied platforms

With this proposed architecture as a roadmap, the focus must now turn to the specific research priorities and fundamental questions that need to be addressed to make this framework a reality.

VIII.FUTURE DIRECTIONS AND OPEN QUESTIONS

8.1 Technical Research Priorities

• **Enhanced Embodied Cognition**: Pushing toward AGI, recent frameworks like **FaGeL** integrate LLM reasoning with rich multimodal sensory inputs—wearable fabrics and ambient sensors—to enable agents to autonomously generate tasks, refine behavior via implicit human feedback, and visualize token-level alignment for interpretability ([arxiv.org][22]). Complementarily, vision-language-action (VLA) agents fine-tune multimodal LLMs with online reinforcement learning, allowing robots not only to act but to strategically ask clarifying questions in ambiguous household scenarios, boosting task performance by up to 40% over zero-shot baselines ([arxiv.org][23]).

• **Advanced Planning and Strategy**: To maintain coherent long-term objectives, the **LIET** paradigm ("Learn as Individuals, Evolve as a Team") equips multi-agent LLMs with both local utility functions for individual decision-making and a shared cooperation knowledge list for team-level strategy, yielding superior performance on benchmarks like Communicative Watch-And-Help and ThreeD-World Multi-Agent Transport ([arxiv.org][24]). Industry forecasts also highlight the arrival of true agentic workflows in 2025—step-by-step reasoning and dynamic subtask reallocation will empower agents to autonomously manage end-to-end processes in business, moving AGI from theory toward practice ([reuters.com][25]).

• **Meta-Learning and Adaptation:** Recursive self-improvement frameworks such as **STOP** ("Self-optimization Through Program Optimization") leverage LLMs to iteratively rewrite and enhance their own code, demonstrating early steps toward systems that can "learn how to learn" without external retraining ([en.wikipedia.org][5]). Meanwhile, evolutionary coding agents like **AlphaEvolve** use LLM-driven mutation and selection loops to discover novel algorithms, showcasing cross-domain transfer and continual adaptation that are essential for scalable AGI architectures ([en.wikipedia.org][26]).

• **Robustness and Reliability:** Ensuring trustworthy AGI requires embedding verifiable reward signals and uncertainty quantification directly into the learning loop. Approaches that fine-tune LLMs with RL-generated rewards—exemplified by Ask-to-Act agents—significantly reduce hallucinations by steering exploration toward behaviors that consistently satisfy verifiable task criteria ([arxiv.org][23]). Ongoing research into confidence calibration and multimodal verification promises to bolster reliability across diverse, real-world operating conditions.

8.2 Safety and Alignment Challenges

- **Value Alignment for Autonomous Agents:** In an agentic paradigm, each agent maintains its own goals and subgoals, making coherent value alignment harder. Research must develop decentralized preference elicitation—where agents negotiate or vote on high-level human values—and robust reward-shaping techniques that propagate ethical constraints through hierarchical policies to every decision node
- **Control and Oversight of Distributed Agents:** Agentic AGI consists of many interacting sub-agents, each potentially acting independently. Ensuring human-in-the-loop oversight requires interpretable policy representations (e.g., symbolic summaries of agent plans), real-time monitoring dashboards that visualize multi-agent workflows, and "circuit breakers" capable of safely pausing or rewiring agentic submodules without collapsing the overall system.
- **Multi-Agent Safety:** Agentic frameworks enable emergent coordination, but also risk coordination failures or cascading errors. Research must explore formal methods for verifying safe interaction protocols, design redundancy mechanisms (so that one agent's misstep can be corrected by peers), and develop runtime anomaly detectors that flag unsafe emergent plans before they execute..

8.3 Fundamental Open Questions

- **Understanding vs. Policy Correlation:** Agentic AGI agents learn to map states to actions via intertwined language reasoning and reinforcement policies. Do these policies reflect genuine world models and causal reasoning, or merely high-dimensional pattern correlations mediated through conversation prompts? Unpacking whether agentic submodules possess true conceptual understanding remains an open challenge.
- **Consciousness and Experience:** Do sufficiently sophisticated AI agents exhibit forms of consciousness, subjective experience, or phenomenal awareness? While difficult to answer definitively, this question affects how we design. interact with, and make ethical decisions about advanced AI systems.
- **Scalability of Current Approaches:** Can the current LLM-based approach to AGI scale to truly general intelligence, or will fundamental architectural changes be required? Understanding the limitations and potential of current approaches is crucial for long-term research planning.
- **Social and Economic Integration:** How will AGI systems integrate with human society, economic systems, and social structures? Research must address the broader implications of AGI deployment beyond technical capabilities.

8.4 Interdisciplinary Research Needs

- **Cognitive Science Integration:** Understanding human cognition, learning mechanisms, and intelligence can inform the design of more effective AGI architectures. Collaboration between AI researchers and cognitive scientists is essential for developing human-like general intelligence.

- **Philosophy and Ethics:** Questions about consciousness, moral agency, and the nature of intelligence require philosophical investigation alongside technical development. Ethical frameworks for AGI development and deployment must be developed through interdisciplinary collaboration.

IX.CHALLENGES AND LIMITATIONS

9.1 Technical Challenges

- **Computational Requirements:** Current LLM-based systems require enormous computational resources, potentially limiting the scalability and accessibility of AGI systems. Research into more efficient architectures, compression techniques, and distributed computing approaches is necessary.
- **Data Requirements:** Training advanced LLMs requires massive datasets, raising questions about data availability, quality, and privacy. Future AGI systems must be able to learn from more limited and diverse data sources.
- **Integration Complexity:** Combining multiple agents, modalities, and capabilities creates significant integration challenges. Managing the complexity of large-scale multi-agent systems while maintaining performance and reliability is a major technical hurdle.

9.2 Safety and Ethical Concerns

- **Unintended Consequences:** As AGI systems become more capable, the potential for unintended and harmful consequences increases. Robust safety mechanisms, testing procedures, and gradual deployment strategies are essential.
- **Misuse Potential:** Advanced AGI capabilities could be misused for harmful purposes, including surveillance, manipulation, or autonomous weapons. Developing appropriate governance frameworks and access controls is crucial.
- **Bias and Discrimination:** LLMs inherit biases from their training data, potentially perpetuating or amplifying social inequalities. Addressing bias in AGI systems requires ongoing research and careful system design.

9.3 Societal and Economic Implications

- **Employment Disruption:** AGI systems capable of performing most human cognitive tasks could lead to massive unemployment and economic disruption. Society must prepare for these changes through education, policy, and economic restructuring.
- **Power Concentration:** Advanced AGI capabilities could concentrate power among those who control these systems, potentially exacerbating inequality and undermining democratic institutions
- **Social Sciences:** Understanding the societal impact of AGI requires input from economists, sociologists, political scientists, and other social science disciplines. Research must address employment effects, power dynamics, and social transformation.
- **Neuroscience:** Insights from neuroscience about brain function, learning mechanisms, and neural architecture can inform the development of more effective AI systems. Brain-inspired approaches may provide alternative pathways to AGI.
- **Human Agency**: As AGI systems become more capable, there is a risk that humans may become overly dependent on AI for decision-making, potentially diminishing human agency and autonomy.

Navigating these profound challenges requires a clear, strategic, and collaborative approach. In conclusion, the path forward must synthesize the lessons learned, leverage the most promising architectural frameworks, and maintain a steadfast commitment to safety and human values

X.CONCLUSION

Achieving Artificial General Intelligence (AGI) demands more than scaling isolated LLMs—it requires a structured, collaborative, and agentic approach. By integrating LLMs into multi-agent systems, we unlock emergent capabilities that surpass the limits of single models. Agentic frameworks enable specialization, memory, planning, coordination, and real-time interaction—core elements essential for general intelligence.

Evidence from recent systems like ChatDev, Devin, and Google's SayCan demonstrates that distributed agents can collaborate to solve complex tasks more effectively than monolithic models. While technical and ethical challenges remain, the agentic paradigm offers built-in modularity, interpretability, and layered safety checks that are difficult to implement in standalone LLMs.

Ultimately, the agentic approach provides a scalable, controllable, and socially aligned pathway to AGI—combining the strengths of LLMs with the structure needed for real-world autonomy and collaboration. It is not just a better method; it is the most viable strategy for building AGI that is safe, robust, and beneficial to humanity.

References

1. Wei, J., et al. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in Neural Information Processing Systems, 35, 24824-24837.
2. Park, J. S., et al. (2023). Generative agents: Interactive simulacra of human behavior. Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, 1-22.
3. Schick, T., et al. (2023). Toolformer: Language models can teach themselves to use tools. arXiv preprint arXiv:2302.04761.
4. Yao, S., et al. (2022). ReAct: Synergizing reasoning and acting in language models. arXiv preprint arXiv:2210.03629.
5. Bansal, T., Kamar, M., Verma, M., Sweihahn, K., & Gupta, K. (2020). Hierarchical Multi-agent Reinforcement Learning for Semi-cooperative Communication Games. Proceedings of the 37th International Conference on Machine Learning, 667-677.
6. Foerster, J., Foerster, J., Kapoor, A., & Schmidhuber, J. (2016). Learning to Cooperate with Hierarchy and Delegation. Proceedings of the 2016 International Conference on Autonomous Agents and Multiagent Systems, 931-939.

7. Hao, J., Yao, H., Liang, P., & Sun, M. (2022). Cooperative Multi-Agent Reinforcement Learning for Team Formation with Unknown Dynamics. Proceedings of the 36th International Conference on Machine Learning, 5073-5083.
8. Liu, Y., Li, M., Wu, Y., Li, S., Xu, L., & Zhu, S. (2023). A Survey of Multi-Agent Communication: From Theory to Practice. arXiv preprint arXiv:2302.11223.
9. Bai, Y., et al. (2022). Constitutional AI: Harmlessness from AI feedback. arXiv preprint arXiv:2212.08073.
10. Bommasani, R., et al. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.
11. Awesome AI Agents Repository. (2024). Retrieved from https://github.com/e2b-dev/awesome-ai-agents
12. Bansal, T., Kamar, M., Verma, M., Sweihahn, K., & Gupta, K. (2020). Hierarchical Multi-agent Reinforcement Learning for Semi-cooperative Communication Games. Proceedings of the 37th International Conference on Machine Learning: https://arxiv.org/abs/2004.07228 (pp. 667-677). PMLR.
13. Foerster, J., Foerster, J., Kapoor, A., & Schmidhuber, J. (2016). Learning to Cooperate with Hierarchy and Delegation. Proceedings of the 2016 International Conference on Autonomous Agents and Multiagent Systems: https://arxiv.org/abs/1602.01724 (pp. 931-939). International Foundation for Autonomous Agents and Multiagent Systems.
14. Hao, J., Yao, H., Liang, P., & Sun, M. (2022). Cooperative Multi-Agent Reinforcement Learning for Team Formation with Unknown Dynamics. Proceedings of the 36th International Conference on Machine Learning: https://arxiv.org/abs/2203.16783 (pp. 5073-5083). PMLR.
15. Lewis, M., Okande, D., & Hernandez-Gardiol, P. (2016). IRL in Multiagent Systems: A Survey. Autonomous Agents and Multi-Agent Systems: [invalid URL removed] 20(3), 685-728. Springer Science+Business Media.
16. Liu, Y., Li, M., Wu, Y., Li, S., Xu, L., & Zhu, S. (2023). A Survey of Multi-Agent Communication: From Theory to Practice. arXiv preprint arXiv:2302.11223: https://arxiv.org/abs/2302.11223.
17. Otte, K., Ray, A., & Silver, D. (2016). Learning to Play Games with Imperfect Information. Proceedings of the 30th International Conference on Neural Information Processing Systems: https://arxiv.org/abs/1603.01121 (pp. 3958-3966). Curran Associates, Inc.
18. Rashkin, H., Choi, E., & Jha, Y. (2018). Toward An Agenda for Deontology in Natural Language Processing. Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency: https://arxiv.org/abs/1806.08770 (pp. 294-305). Association for Computing Machinery.
19. Schulman, J., Lehman, S., & Leibo, J. Z. (2015). Intrinsically Motivated Reinforcement Learning for Sequential Decision-Making. Proceedings of the 32nd International Conference on Machine Learning: https://arxiv.org/abs/1506.05869 (pp. 1714-1722). PMLR.
20. Serban, C., Bohnert, A., Courville, A., Lowe, R., Morin, Y., & Wu, S. (2017). A Comprehensive Survey of Deep Reinforcement Learning. arXiv preprint arXiv:1701.07281: https://arxiv.org/abs/1701.07281.
21. Fig 2 and 3 source : https://github.com/e2b-dev/awesome-ai-agents
22. https://arxiv.org/abs/2412.20297?utm_source=chatgpt.com "FaGeL: Fabric LLMs Agent empowered Embodied Intelligence Evolution with Autonomous Human-Machine Collaboration"
23. https://arxiv.org/abs/2504.00907?utm_source=chatgpt.com "Grounding Multimodal LLMs to Embodied Agents that Ask for Help with Reinforcement Learning"
24. https://arxiv.org/abs/2506.07232?utm_source=chatgpt.com "Learn as Individuals, Evolve as a Team: Multi-agent LLMs Adaptation in Embodied Environments"
25. https://www.reuters.com/technology/artificial-intelligence/autonomous-agents-profitability-dominate-ai-agenda-2025-executives-forecast-2024-12-12/?utm_source=chatgpt.com "Autonomous agents and profitability to dominate AI agenda in 2025, executives forecast"
26. https://en.wikipedia.org/wiki/Recursive_self-improvement?utm_source=chatgpt.com "Recursive self-improvement"