AGI via Multi-Agent Systems: Towards a Scalable and Adaptive Intelligence Model

Malhar P. Ubhe
NBNSTIC Pune
INDIA

Rahul M. Samant
NBNSTIC Pune
INDIA

ABSTRACT
Artificial General Intelligence (AGI), one of the most crucial domains of Artificial Intelligence (AI), which aims to develop systems that are capable of performing broad-scale cognitive tasks which require human intelligence and tasks that require more than narrow intelligence. The major challenge is to identify such tasks that could be termed as AGI tasks and determining the techniques required to attain general intelligence. This review paper clearly defines the category of AGI tasks and explores the existing techniques for AGI development. A novel framework is proposed using tools like Autogen, LangChain, and Phidata to develop a multi-agentic workflow for performing AGI tasks that would define a future path towards AGI development.

General Terms
Artificial General Intelligence, Artificial Intelligence, Multi-Agent Framework, AGI Tasks, Cognitive Architecture.

Keywords
Artificial Intelligence, Agentic Collaboration Workflow, Agents and Large Language Models, Multi-Agent Framework.

1. INTRODUCTION
The next goal in AI research is to achieve Artificial General Intelligence (AGI), which is an advanced version of Artificial Narrow Intelligence (ANI). In the broader view AGI is the ability to think, act and learn like humans and execute the task with absolute perfection. In other words, AGI systems must show human-like behavior of learning new things, rectifying errors and achieving perfection while executing the tasks. This type of intelligence is the key factor which differentiate ANI system from AGI system and similarly the tasks that AGI systems can perform. These are the complex AGI tasks which ANI systems fail to execute satisfactorily. These tasks require high cognitive power for a system to execute them. A perfect example of this complex AGI task is in which emotional intelligence is required, most of the ANI systems present today cannot experience emotions or subjective feelings, their responses are totally based upon the datasets and are analytical [1]. In general, AGI tasks which require complex multi-tasking across different domains, understanding things beyond text i.e. through visual interpretation. Without a well-defined set of tasks to be AGI tasks, it becomes difficult to assess the progress in the development of AGI systems and also the comparison between different approaches becomes incomprehensible. This review research paper addresses two major research questions:
(a) What tasks can be defined as AGI tasks?
(b) How can collaboration of LLMs and multi-agent framework lead towards AGI development?
So, to address these questions, various existing techniques of AGI development were analyzed, leading to the proposal of a refined framework that outlines essential task properties and evaluation criteria for the development of AGI systems. This refined approach towards AGI, consists of integrating open-source large language models like Phi3:medium, DeepSeek R1 versions-14B & 32B, mistral small and other such models on Ollama, with multi-agent workflows like Autogen, Phidata, LangChain and other RAG applications.

2. LITERATURE REVIEW
After the release of ChatGPT by OpenAI, the entire world's attention was drawn towards AI and rapid research work was initiated for exploring the various methodologies which would eventually lead to Artificial General Intelligence (AGI). As a result, AGI has gained significant attention due to its potential to create autonomous, human-like intelligence capable of general problem-solving. This section of the paper reviews existing AGI approaches, highlighting key developments and research gaps in the work done up till now. Before AGI discussions must be done on the significance of Large Language Models (LLMs) and their impact in AGI development. Though LLMs showcase impressive standards in linguistic fluency and pattern recognition, they lack the essential attributes that a true AGI system must persist which includes logical reasoning and autonomous decision making. The study of Ben Goertzel [2] on LLMs involves that though today's LLMs are impressive in carrying out various tasks, they rely on statistical data rather than deep understanding which creates an obstacle for achieving AGI as it demands for continuous learning and real-world interactions. The two major promising architectural pathways for AGI developments are monolithic models and multi-agent systems. In monolithic models it is believed that as these models are large and rely their scaling on deep learning models, they will eventually exhibit general intelligence [3]. However, the second approach of multi-agent systems is more promising for achieving general intelligence as it involves multiple domain specific agents for solving complex tasks [4]. Another approach towards AGI is to strengthen the cognitive architectures like symbolic AI and connectionist models. The study suggests that though symbolic AI is efficient in structured reasoning, but it lacks in adaptability, while connectionist models are effective for pattern recognition but struggle with logic and explanation. A hybrid mode which involves both these cognitive approaches with multi-agent cooperation is more efficient [5]. The assumption that scaling laws for neural networks will give a pathway towards AGI is discarded as though performance improves with large data and parameters, but the fundamental limitation of the models still persists. Thus, it can be stated as scaling alone does not address key challenges like generalization beyond training data, causal reasoning, and real-world adaptability [6]. A practical approach in multi-agent framework is seen in Microsoft's Autogen, a multi-agent conversation model, it is an open-source framework designed to create multi-agent AI systems that use Large Language Models (LLMs) for carrying out various complex tasks [5]. Autogen has shown a pathway towards AGI by its

conversational approach between multiple agents in a workspace. It also allows for human intervention in these agentic conversations when necessary, making it more reliable and adaptable as per the human needs. Autogen supports different conversation patterns like sequential, nested, and group chats to structure complex AI workflows. This agentic conversation is controlled by a manager agent whose goal is to coordinate the talks between multiple domain specific agents for executing complex tasks [7]. The approach stated is that the AGI development must focus on integrated models rather than isolated AI techniques. The hybrid approach discussed, where symbolic reasoning, probabilistic inference, and deep learning are combined to achieve general intelligence. The evolution of chatbots has also led to enhancements in ANI technologies, though AGI can't be achieved through existing chatbot technologies as chatbots can mimic human conversation but fail to understand human emotions, reasoning and general problem-solving capabilities [8]. According to Peter Voss, self-learning is one of the most critical features for a system to be declared AGI. The system must be adaptive, goal-directed and must have the capability to learn from previous as well as real-time novel scenarios. The general intelligent system is not something which has a lot of knowledge and skill set but instead it is a system which is able to acquire the knowledge from novel interactions and improve these capabilities and apply them appropriately [9]. The major characteristics of the AGI system is that it must be able to independently learn accurately and provide goal-oriented responses to input text from the user [10]. Existing research studies in AGI development have explored various methodologies from cognitive architectures and large language models to monolithic models and deep learning. Despite these advancements, the current methodologies struggle to meet the characteristics of AGI systems to have an autonomous learning approach, adaptability, common-sense understanding and decision making [11]. The approach towards AGI development is based on the multi-agent framework of Microsoft's Autogen and Large Language model through Ollama such as the Phi3:medium model. This study proposes an enhanced multi-agent system for AGI, incorporating adaptive learning techniques to improve generalization across varied tasks. The following section outlines the methodology used to develop and evaluate this proposed multi-agent framework.

3. PROPOSED SYSTEM

Microsoft's Autogen is a multi-agent framework model that uses various Large Language models and multiple agents for solving complex tasks in a very efficient and effective manner. Autogen allows us to impart skills to various agents that are deployed in the workspace playground for carrying out various tasks given by the user. Autogen allows for creating various domain specific as well as general workspace for executing complex tasks [5]. The working mechanism of Autogen is in a collaborative manner where just as like a human team methodology, there is a team of multiple AI agents with specific skills connected to pre-trained LLMs. This team goes forward with a group chat of AI agents where-in initially they discuss how to solve the complex task and then creates a plan and every agent acts accordingly [12]. The complex problem is broken down into various sub tasks and allotted to specific agents for execution, an agent named group chat manager is always in contact with all the agents working in the workspace playground for carrying out the tasks and provides necessary resources to the agents for task execution. There is also a provision of human intervention whenever required, as the agents change their approach to problem-solving of complex tasks as per the human inputs [13]. The results of tasks allotted to each agent are summarized together and presented to the user as the final output of the complex task.

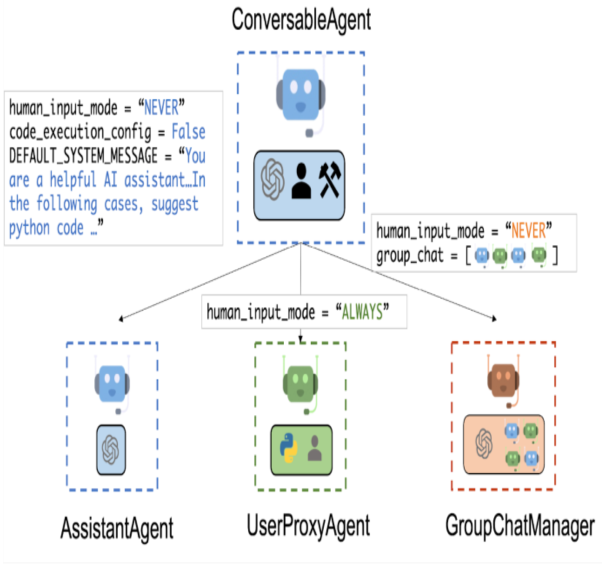

Fig 1: Architecture of multiagent framework of Microsoft's Autogen. Screenshot taken from the official Microsoft Autogen GitHub repository:
https://microsoft.github.io/autogen/0.2/docs/Use-Cases/agent_chat/

LLMs in this framework play a crucial role at various stages just like an actor on the stage, such as, for understanding natural languages for interpreting the user queries and generating natural language responses for these queries and also for holding talks between agents for creating a human-like discussions [14]. LLMs also act as task decomposer, i.e., the LLM behind planner agent is responsible for breaking large and complex tasks into more manageable and smaller subtasks and assigning them to specific specialized agents which handle a particular part in complex task-execution.

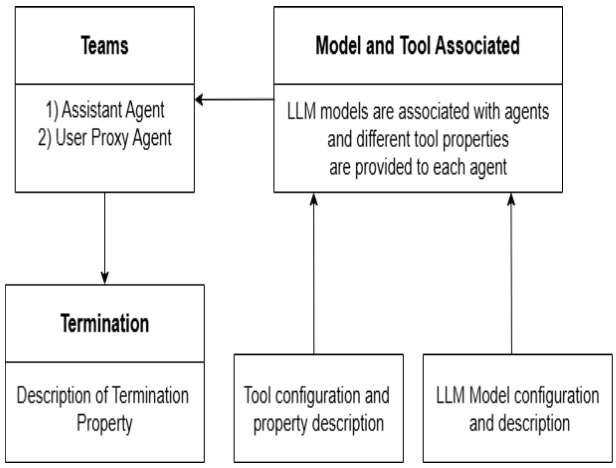

Fig 2: General Architecture of Autogen and its underlying workflow mechanism.

LLMs allows agents to retrieve and study the information in the documentation, APIs, external databases and make more informed and correct decisions [15]. LLMs also help in identifying inconsistencies in the work done by specific agents and handle these, make proper modification to treat these inconsistencies and generate more efficient and reliable responses[16].

An important feature of Autogen is that it allows to use different LLM models like OpenAI's GPT- 4 Family, GPT - 3.5 Family and many other open source Ollama LLM models with different agents for solving the tasks in an efficient manner. With this approach, the focus will be on integrating a Phi3:medium LLM model, DeepSeek R1 model and GPT- 4 family models with the agents in the Autogen's workspace.

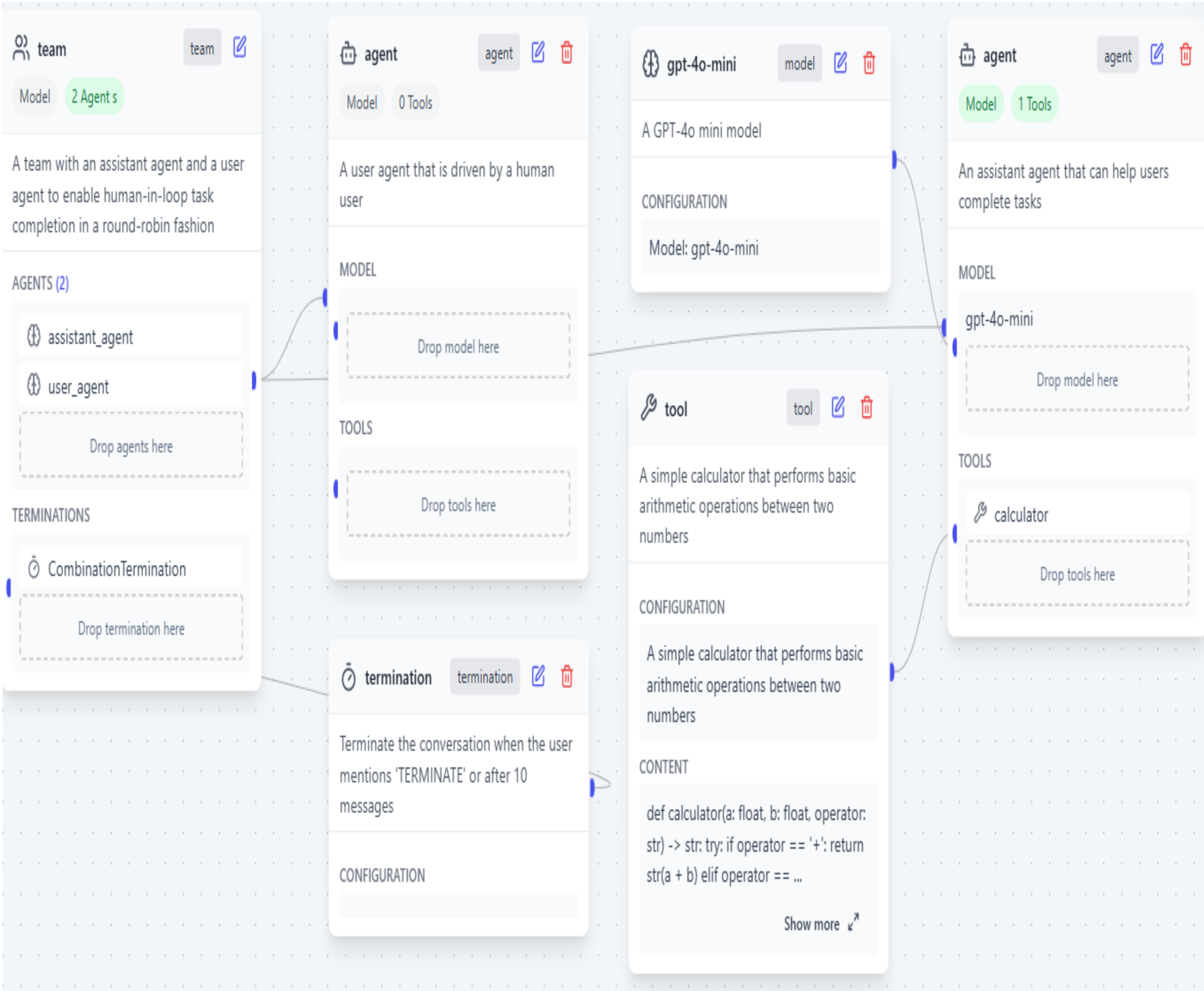

Fig 3: General Workflow of Task Execution Environment of Autogen. Screenshot from local execution of Microsoft's Autogen interface (ver. 0.4), demonstrating multi-agent interaction. Microsoft Autogen GitHub repository:
https://microsoft.github.io/autogen/stable//user-guide/agentchat-user-guide/installation.html

Following are the few tasks that Autogen has performed with great efficiency:

a. Task 1: Extracting text, links, images from a website when the web URL link is provided.

A skill for fetching information from a web page can be created in Autogen and then integrated with specific agents in Autogen's workspace. The main goal of these agents is to extract text, hyperlinks and image-source links as specified by the user. The GPT - 4 Family LLM model was used for real time processing for execution of the task.

b. Task 2: Code Generation and Error Handling.

Autogen's agents are very efficient in generating codes of Python, Java and other programming languages as per the specified directions of the user. The LLM Model used here is Ollama's DeepSeek R1 model. A Coder agent named agent has been given the task for generating codes based on the task provider and an error detection agent is used for debugging and suggesting improvements to be done in the code. Based on these suggestions the Coder agent modifies the code based as per the suggestions.

Human responses are also taken into consideration and specific modifications are done in the code.

c. Task 3: Parsing Large information in the documents and summarizing this information and answering query questions.

A document-reader agent is used to read the PDFs, Word files and other large documents. The LLM model used here is Ollama's Phi3:medium model. A summarization agent is responsible for summarizing the entire text content and providing the results to the user. The Q&A agent is the one who interacts with the user and answers the queries presented by the user by extracting relevant information from the document.

d. Task 4: YouTube URL to Blog and Tweet Thread Generation.

A transcription agent is responsible for extracting the transcripts of the YouTube video from the URL provided using the YouTube API Keys which the coder agent on its own adds in the code if required. A Blog Generation agent converts these transcripts into a structured blog post and is presented to the user. A Tweet Thread Generation agent converts this blog Post into multiple tweets and presents the results to the user.

e. Task 5: Surfing real-time news on the internet and giving updates.

The search and retrieval agent surfs the internet in search of the relevant news articles as per the user requirements using search engines, API and web scraping to find latest news on specific topics from reliable sources. The Summarizing agent extracts relevant information and creates concise reports by handling misinformation and then delivery agent presents this generated news reports to the user.

f. Task 6: Stock Market Analysis.

The data retrieval agent gathers all the historical and real-time data related to the specified stock which includes its trading volume, historical performance from various financial data APIs. The news analysis agent then processes all the latest financial news articles, reports for understanding market sentiments and based on these insights the technical analysis agent applies various technical methods and other indicators to detect trends and opportunities in trading.

All these above tasks are performed successfully by using the Multi-Agent Collaboration concept of AI by Autogen which would eventually lead us a pathway towards achieving AGI.

4. COMPARISON & DISCUSSIONS

4.1 Task Orchestration

Autogen's task orchestration is mainly based on its multi-agent collaboration behavior, where multiple AI agents work together as a team for executing various tasks in a structured manner with a flexible approach.

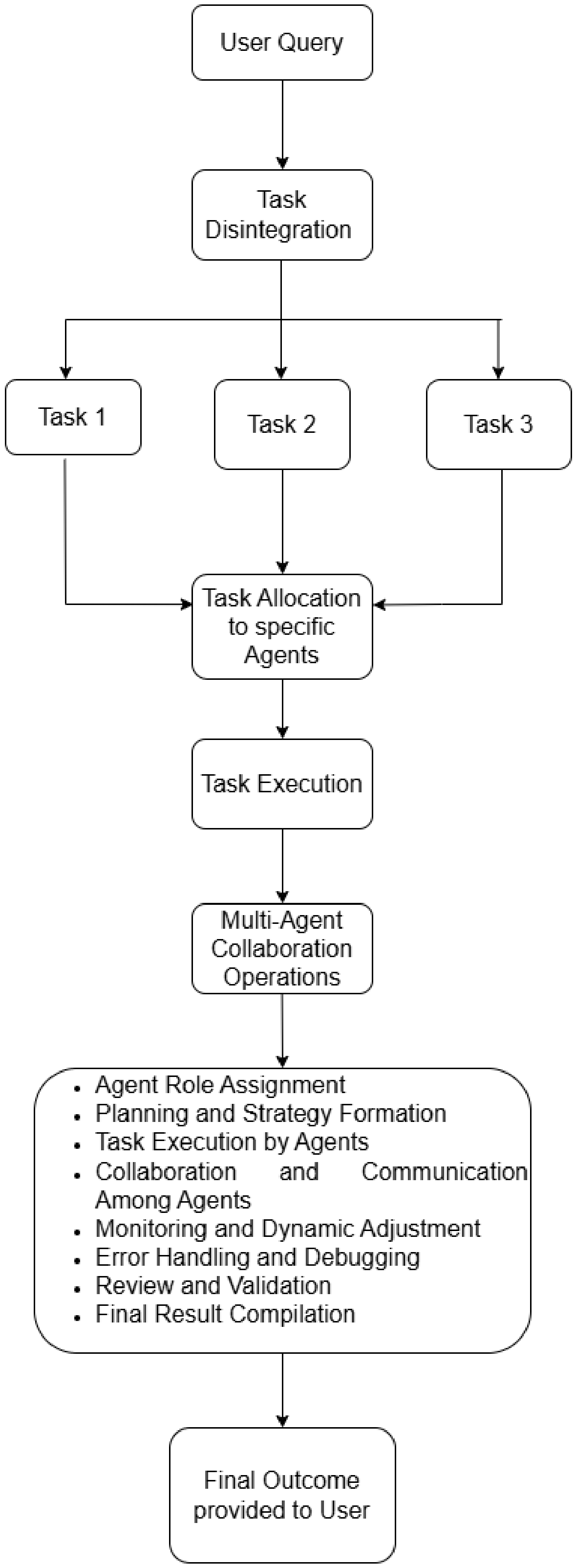

Fig 3: Autogen's General Task Orchestration Flow while executing any task.

When compared with other traditional models where the entire working of systems used to be based on predefined set of rules, Autogen breaks these traditional beliefs and introduces a flexible approach of collaboration of different skilled agents [17]. It allows agents to dynamically interact, negotiate their needs and bring out success for their assigned tasks. These agents work in various roles such as planner, manager, coder, debugger, reviewer, etc. ensuring that all the tasks are handled with great efficiency and with minimum human intervention. Autogen's task orchestration depicts an important feature of AGI systems i.e. adaptive workflow execution [18]. In general, where traditional AI systems follow a fixed approach for solving a problem, Autogen's agents continuously monitor their progress by updating the results to the other agents within the workflow and adjust their approach of problem solving based on real time feedback [19]. This approach allows it to handle errors, re-evaluation of responses and dynamic modifications in tasks making it highly effective for complex decision making. Autogen also enhances its capabilities by integrating with other tools and APIs making it more reliable in diverse and multi-step process tasks.

4.2 Other Frameworks

Autogen is based on a promising AGI architecture - The Multi-Agent Systems. This agentic workflow mechanism is also implemented in various other tools like Phidata, LangChain and Hugging Face Transformers. These Agentic AI tools have various features making them more reliable and adaptable [20]. Other AI framework models like LangChain, Phidata and N8N follow a similar working pattern but differ in its approach for executing complex problems. LangChain is a modular framework which is totally based on designing applications by using Large Language Models [21]. It acts as a connector between AI applications and LLMs as its main goal is to simplify the integration of LLMs into various complex workflows by introducing services that help enhance memory handling, provide pre-defined chains and API connectors [22]. As compared to Autogen, LangChain is more focused on structured prompt execution along with human interventions while making adjustments in the workflows to obtain the desired output. Phidata's main motto is to deal with data driven tasks, wherein it carries out ETL (Extract, Transform , Load) tasks and AI powered analytics and data pipeline management with great efficiency [23]. It provides a Python backed workflow automation mechanism which allows users to handle large scale data transformations with ease. Unlike Autogen it doesn't involve multi-agent framework and AI driven interactions but for AI assisted analytics [24]. The collaboration of the multi-agent framework of Autogen and the data focused automation framework of Phidata will make the process of executing complex tasks involving processing huge data sets easier by ensuring seamless AI driven insights for business intelligence. N8N is a no-code automation framework whose main task is to construct a workflow by API integration for specializing repetitive business operations, collaborating multiple third-party applications and managing API calls in an effective manner [25]. In general, all the above specified frameworks focus on automation and AI driven methodologies for solving varied tasks, Microsoft's Autogen stands out in all these tools as it is an ideal intelligent multi-agent collaboration framework which has the potential to solve any complex task in a structured way. Though collaboration of all these tools with Autogen will definitely lead to a more reliable framework which will be able to execute any real-world complex task. Some comparison of Autogen's Multi-Agent framework and other frameworks is given in the below table,

Table 1. Comparison of Autogen and its Multi-Agent framework with other existing frameworks.

Aspect	Autogen (Multi-Agent AI)	LLM Standalone	Traditional RPA	Monolithic Architecture
Task Complexity	Can handle complex tasks with agentic collaboration.	It struggles handling complex tasks due to single prompt execution.	It fails for handling complex tasks as it is based on rules.	Efficient in handling complex tasks but lacks modular flexibility.
Error Handling	Display of dynamic debugging and instant error handling	There is no built-in self-correction mechanism.	Manual debugging is required.	Errors affect the entire system, and it becomes difficult to isolate issues.
Scalability	It is modular and scalable with multiple agents.	It is difficult to scale beyond one module.	It can be scaled up to some extent but requires updating the rules.	Scaling requires increasing hardware resources.
Adaptability	It can adapt to novel scenarios dynamically.	It cannot showcase adaptability as output is based on the input prompts provided by the user.	It is rigid and totally based on rules.	It showcases limited adaptability as it requires entire system updates.
Execution Speed	It follows fast parallel execution with multiple agents.	It follows sequential processing of tasks and is single threaded and hence comparatively slower	It can be fast but totally dependent on predefined rules.	It can be optimized for high-speed execution but may face bottlenecks due to centralized design.
Self-Learning Capabilities	It can learn from feedback and while task execution and reinforcement learning.	No self-learning feature.	It does not have the feature of self-learning.	It can have built-in learning capabilities, but it will have to retrain the entire system.

5. CONCLUSION & FUTURE DIRECTIONS

The development of Artificial General Intelligence is one of the most ambitious and complex challenges in Artificial Intelligence research. This review has highlighted the key characteristics of AGI development which includes architectural innovation, Agentic workflow mechanism, self-learning capabilities and adaptability. The distinction between standalone large language models, traditional rule-based automation and the multi-agent framework architecture has also been highlighted. As compared to other architectures the multi-agent framework proves to be a promising mechanism for achieving AGI as it showcases adaptability, reliability, fault tolerance and has a collaborative nature making its behavior closer to that of humans while solving any problem. A key finding from the research review is that, at present, there is no specific singular framework approach for achieving AGI. Instead, a hybrid approach where different cognitive AI workflow mechanisms are involved with each other is required for achieving AGI. Though multi-agent framework is more efficient, if it is integrated with other domain specific mechanisms, it will enhance the cognitive capabilities of the system and hence will lead towards a step ahead in achieving AGI. One of the most important gaps in AGI research is, there is no well-defined specific set of tasks for which we could call them as AGI tasks. There is a need for a new evaluation benchmark which assesses the system's ability to carry out autonomous, reasoning tasks and is a system capable of maintaining a self-learning approach. Another major limitation that existing AI systems face is task generalization. Today's ANI systems are capable of dealing with domain specific tasks but when a particular problem involves interdisciplinary collaboration and transfer of knowledge across domains these systems fail. Hence there is a need for carrying out extensive research in task generalization along with adaptive learning. Thus, there is a need for future AI research to focus on overcoming these gaps and enhancing AI power as enhancing the existing multi-agent workflow will significantly improve the decision making, task execution and adaptive learning. The focus must be on strengthening human-AI collaboration as AI must not replace human intelligence but instead augment human intelligence. In conclusion AGI development requires an interdisciplinary hybrid approach of integrating various advancements in AI, cognitive workflows and AI architectures. By addressing all these challenges in future research, AGI can evolve into a secure intelligence which will eventually lead to human progress by focusing on human values and ethical considerations.

6. REFERENCES

[1] Cassio Pennachin and Ben Goertzel, 2007. "Contemporary Approaches to Artificial General Intelligence", Artificial general intelligence - Springer Berlin Heidelberg, pp. 1-30.

[2] Ben Goertzel, 2023. "Generative AI vs. AGI: The cognitive strengths and weaknesses of modern LLMs", ArXiv preprint arXiv:2309.10371.

[3] Nanyi Fei, Zhiwu Lu, et al, 2022. "Towards artificial general intelligence via a multimodal foundation model", Nature Communications, 13/1, pp. 3094.

[4] Fernando Diaz, and Michael Madaio, 2024. "Scaling laws do not scale", Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society. 7, pp. 341-357.

[5] Qingyun Wu, Gagan Bansal, et al, 2023. "AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation", arXiv preprint arXiv:2308.08155, 3/4.

[6] Rahul M. Samant, et al, 2013. "A Study on Feature Selection Methods in Medical Decision Support Systems", International Journal of Engineering Research & Technology (IJERT), 2/11, pp. 615-619.

[7] Ben Goertzel, " Artificial General Intelligence : Concept, State of the art, and Future prospects", Journal of Artificial General Intelligence, 5/1, 2014.

[8] J. Grudin and Richard Jacques, 2019. "Chatbots, Humbots, and The Quest for Artificial General Intelligence", Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1-11.

[9] Peter Voss, 2007. "Essentials of General Intelligence: The Direct Path to Artificial General Intelligence", Springer Berlin Heidelberg, pp. 131–157.

[10] Sébastien Bubeck, Varun Chadrasekaran, et al, 2023. "Sparks of Artificial General Intelligence: Early Experiments with GPT-4 ", ArXiv, 2303.

[11] Rahul Samant and Srikantha Rao, 2013. "A study on Comparative Performance of SVM Classifier Models with Kernel Functions in Prediction of Hypertension", International Journal of Computer Science and Information Technologies, 4/6, pp. 818-821.

[12] Minqi Jiang, Tim Rocktäschel, Edward Grefenstette, 2023. "General intelligence requires rethinking exploration", Royal Society Open Science, 10/6, pp. 230539.

[13] Lin Zhao, Lu Zhang, Zihao Wu, Yuzhong Chen, Haixing Dai, Xiaowei Yu, et al, 2023. " When brain inspired AI meets AGI ", Meta-Radiology - Elsevier, 1/1, pp. 100005.

[14] James Babcock, János Kramár and Roman Yampolskiy, 2016. "The AGI containment problem", Artificial General Intelligence: 9th International Conference, AGI 2016, New York, NY, USA, July 16-19, 2016, Proceedings 9, Springer International Publishing, pp. 53-63.

[15] Andrzej Cichocki and Alexander P Kuleshov, 2021. "Future trends for Human-AI collaboration: A Comprehensive Taxonomy of AI/AGI using Multiple Intelligences and Learning Styles", Computational Intelligence and Neuroscience - Hindawi, 2021/1, pp. 8893795.

[16] Rahul Samant and Srikantha Rao, 2013. "Evaluation of Artificial Neural Networks in Prediction of Essential Hypertension", International Journal of Computer Applications, 81/12.

[17] T. Everitt, G. Lea and M. Hutter, 2018. " AGI Safety Literature Review ", International Joint Conference on AI.

[18] D. B. Acharya, K. Kuppan and B Divya, 2025. "Agentic AI: Autonomous intelligence for complex goals – A comprehensive survey", IEEE Access.

[19] Erik Miehling, Karthikeyan Natesan Ramamurthy, Kush R Varshney, Matthew Riemer, Djallel Bouneffouf, John T Richards, Amit Dhurandhar, et al, 2025. "Agentic AI Needs a Systems Theory", arXiv preprint arXiv:2503.00237, IBM Research.

[20] Shanmugasundaram Sivakumar, 2024. "Agentic AI in Predictive AIOps: Enhancing IT Autonomy and Performance", International Journal of Scientific Research and Management (IJSRM), 12/11, pp. 1631-1638.

[21] Antonio Lieto, Mehul Bhatt, Alessandro Oltramari, David Vernon, 2018. "The role of cognitive architectures in General Artificial Intelligence", Cognitive Systems Research, Volume - 48, pp. 1-3.

[22] Steve J Bickley and Benno Torgler, 2023. "Cognitive architectures for artificial intelligence ethics", AI & Society, 38/2, pp. 501-519.

[23] Qian Niu, Junyu Liu, et al, 2024. "Large Language Models and Cognitive Science: A Comprehensive Review of Similarities, Differences, and Challenges", arXiv preprint arXiv:2409.02387.

[24] Henry Hexmoor, Johan Lammens, et al, 2025. "Behaviour Based AI, Cognitive Processes, And Emergent Behaviors in Autonomous Agents", WIT Transactions on Information and Communication Technologies, Volume-1.

[25] Teppo Felin, Matthias Holweg, 2024. "Theory Is All You Need: AI, Human Cognition, and Decision Making", Strategy Science, 9/4, pp. 346-371.