AI Agents in Research Automation: A Comprehensive Review

Bhagyashri R. Wankar, Achyut Shinde, Bhavesh Choudhary, Kartik Aswar, Sonali M. Sonavane, Komal R. Jadhav
G H Raisoni College of Engineering & Management, Pune, India
Emails: bhagya.wankar@gmail.com, avshinde018@gmail.com, choudharybhavesh191@gmail.com, kartikaswar03@gmail.com, sonali.sonavane@raisoni.net, komal.jadhav@raisoni.net
Sagar Satish Gaikwad
Vice President, ACTIN Technologies, Pune, India
Email: sagar.gaikwad2829@gmail.com

Abstract—Large language models have, somewhat quietly, pushed Retrieval-Augmented Generation (RAG) into the center of many research-oriented applications. The underlying idea is not particularly complicated. Instead of relying only on what a model absorbed during training, a RAG system reaches outward. It retrieves relevant information from external sources and then builds its response around that material. In theory, that sounds sensible enough. However, in practice, many conventional RAG setups still depend on relatively static knowledge bases. These repositories can age quickly, especially in domains where information shifts almost weekly. As a result, the system's reasoning can feel narrow. Verification steps are often minimal, and incorporating newly emerging knowledge is, at times, more cumbersome than one might expect.

Because of these limitations, researchers have started looking beyond the traditional retrieval pipeline. Recently, attention has moved toward more agentic forms of AI-driven research systems. The ambition here is a little different. Rather than acting as passive tools that fetch documents and summarize them, these systems are designed to behave more like active research assistants. Work on agent-based reasoning, autonomous research workflows, multi-agent coordination, and self-correcting AI mechanisms hints at a more fluid research environment [?]. In principle, such systems attempt to reproduce some aspects of how human researchers actually work. They plan what to investigate, discover information in real time, examine documents more deeply, compare ideas across multiple sources, and occasionally check their own conclusions through iterative verification.

With that backdrop, this review looks closely at the emerging landscape of agentic research systems. The goal is to identify their strengths, the weaknesses that still persist, and the design patterns that appear to be taking shape. Some of these approaches do suggest meaningful improvements for AI-assisted research. Yet the picture is not entirely settled. Questions around reliability, scalability, and real-world deployment remain open, and perhaps they will continue to be for some time.

Index Terms—Agentic AI, Multi-Agent Architecture, Retrieval-Augmented Generation (RAG), Self-Correcting AI, Large Language Models, Information Retrieval, Document Intelligence.

I. INTRODUCTION

The rapid progress of artificial intelligence, especially large language models (LLMs), has begun to reshape how people search for, interpret, and piece together information. Tasks that once meant spending hours combing through papers, comparing different sources, or extracting useful ideas can now be partly assisted by systems that process large amounts of text in minutes. Researchers, students, and professionals increasingly turn to these tools not just for quick answers but also for that early stage of exploration when a topic is still unfamiliar and the direction of inquiry is not entirely clear.

Early AI-supported research tools, however, were built on fairly rigid retrieval and generation methods. One of the most influential approaches in this space is Retrieval-Augmented Generation (RAG). Put simply, RAG pairs a pretrained language model with an external knowledge source, allowing the system to retrieve relevant material before composing a response. The idea is appealing because it grounds the generated answer in retrieved evidence rather than relying solely on what the model memorized during training. In many straightforward cases, especially fact-oriented queries, the approach works reasonably well. Yet the situation becomes more complicated when a research task demands deeper reasoning or synthesis across multiple documents. Conventional RAG systems, for instance, often struggle when they must connect ideas scattered across several sources or adapt to information that changes rapidly. Long-term planning is limited, cross-source reasoning can feel shallow, and mechanisms for verifying outputs tend to be underdeveloped [1].

Partly because of these limitations, attention has gradually shifted toward what many researchers now call agentic AI systems. The basic motivation is easy enough to grasp. Rather than treating AI as a passive generator that retrieves and summarizes text, these systems attempt to operate more like active research assistants. They emphasize autonomy, goal-directed behavior, and the ability to perform multi-step reasoning while interacting with external tools or environments. In practice, such systems may divide a research objective into smaller tasks, search for information in several rounds, compare evidence from different sources, and adjust their conclusions over time. That process feels oddly familiar. Human researchers rarely reach an answer in one attempt; they explore possibilities, revise assumptions, and sometimes abandon a line of thinking when new information appears. Agentic architectures attempt, in a limited way, to reproduce aspects of that workflow. For this reason, they are increasingly discussed as promising foundations for more sophisticated AI-supported research pipelines [2], [3].

This review examines recent developments in agentic AI for research automation through the conceptual lens of what might be called an agentic research engine. The goal is not simply to catalogue individual tools or models. Instead, the intention is to understand how several emerging ideas connect and where they might be leading. By looking at work on agent-based architectures, autonomous reasoning frameworks, multi-agent collaboration, and self-correcting mechanisms, the discussion attempts to trace patterns that seem to be forming across the literature. Some proposed systems rely on multiple specialized agents that cooperate, each responsible for tasks such as retrieval, reasoning, or evaluation. Others emphasize iterative self-improvement, where models critique and refine their own outputs through structured feedback loops. Taken together, these approaches hint at a gradual movement away from rigid pipeline-based designs toward architectures that are more flexible and adaptive.

At the same time, the field is still finding its footing, and a number of practical challenges remain unresolved. Questions about reliability, computational cost, coordination among multiple agents, and even how one should evaluate the quality of AI-assisted research are still actively debated. Some implementations appear promising in controlled demonstrations but become harder to manage at scale. Therefore, this review considers both the potential and the current limitations of emerging agentic research systems. By examining methodological trends, architectural ideas, and the unresolved challenges that accompany them, the paper aims to clarify how these systems attempt to move beyond the constraints of conventional RAG-based frameworks and move toward AI-supported research processes that are more dependable, adaptable, and transparent [4]–[6].

II. BACKGROUND

Artificial intelligence has moved through several fairly distinct phases over the past few decades, and each phase has carried its own assumptions about how machines should "think." Early systems leaned heavily on rule-based expert frameworks. Knowledge was written down manually, often as long chains of logical rules crafted by specialists who tried to anticipate every possible scenario. In domains like medical diagnosis or industrial troubleshooting, these systems could be impressively precise, but only when the situation matched the rules they had been given. Outside those boundaries, they tended to fail rather abruptly. Later, the field shifted toward data-driven machine learning models. Instead of encoding knowledge explicitly, researchers began training algorithms on large datasets so that the system could learn statistical patterns from examples. Techniques such as decision trees, support vector machines, and neural networks gradually became common tools for solving classification and prediction tasks. More recently, large language models (LLMs) have extended this trajectory in an unexpected way. Trained on massive corpora of text, these models can interpret context, summarize information, and generate coherent language with a fluency that would have seemed difficult to imagine not long ago [7].

Each stage expanded the capabilities of AI systems, although the limitations were rarely subtle. Traditional AI tools were typically designed for narrowly defined tasks operating under predetermined conditions. Once those conditions changed, reliability often dropped. Updating such systems required significant human effort, since new rules or training data had to be manually incorporated. Moreover, these systems rarely demonstrated the kind of autonomy required in complex research settings. Open-ended inquiry is messy by nature. A researcher may explore several sources, revise hypotheses, and reinterpret evidence over time. Conventional AI systems, even fairly advanced machine learning models, were not particularly good at handling that sort of iterative reasoning. They performed well in controlled benchmarks but were often less dependable when applied to real-world knowledge discovery [3].

The arrival of large language models marked a noticeable shift in this progression. Because they can process extended context and generate coherent natural language, LLMs opened the door to applications centered on knowledge exploration and synthesis. Researchers quickly began experimenting with ways to connect these models to external information sources. One development that gained widespread attention was Retrieval-Augmented Generation (RAG). In simple terms, RAG connects a language model to external knowledge repositories such as research articles, structured databases, institutional archives, or even curated web collections [1]. Before generating a response, the model retrieves documents that appear relevant to the query. The retrieved passages are then incorporated into the generation process, allowing the model to produce answers grounded in specific evidence. This approach helps reduce the risk of fabricated statements, often referred to as hallucinations, and can significantly improve factual alignment in domains where accuracy matters [8].

Yet in practice these systems remain somewhat reactive. A typical RAG pipeline retrieves information and produces an answer in a single step. There is usually little opportunity for the system to reconsider its reasoning or to cross-check evidence across multiple documents. If a retrieved source is incomplete or misleading, the generated output may inherit those limitations. Tasks that require deeper analysis, such as comparing conflicting findings from several research papers or synthesizing arguments across an entire body of literature, remain challenging. Planning a sequence of investigative steps, verifying intermediate conclusions, or refining results through iterative reasoning is generally outside the capabilities of most standard RAG implementations [2].

In response to these limitations, a growing body of research has turned toward what is often described as agentic AI. The idea here shifts the focus from passive information retrieval to autonomous decision-making processes. Instead of simply generating text, agentic systems are designed to behave more like problem-solving entities that can set goals, plan actions, interact with external tools, and adapt their behavior based on feedback [3], [9]. Architectures of this kind often include several interacting components. Memory modules can store intermediate findings or contextual knowledge. Reasoning modules help interpret retrieved information and guide subsequent actions. Reflection or self-evaluation mechanisms allow the system to critique its own outputs and attempt improvements. Together, these elements enable more complex workflows that resemble the iterative nature of human research.

The concept becomes even more compelling in multi-agent environments. Rather than relying on a single model, several specialized agents may collaborate within the same system. One agent might focus on retrieving relevant documents, another might analyze and summarize evidence, while a third evaluates the credibility or consistency of the conclusions. Coordination mechanisms then allow these agents to exchange information and refine their outputs collectively. In some experimental systems, this collaborative structure has been shown to improve reasoning depth and reduce certain types of errors, particularly when tackling large or complex research questions [4], [5].

Seen from this perspective, agentic research engines appear as a natural continuation of the broader evolution of AI-driven research tools. They attempt to combine the generative capabilities of LLMs, the grounding benefits of retrieval-based systems, and the autonomy provided by agent architectures. Instead of answering isolated queries, these systems aim to support entire research workflows. That may include discovering relevant literature, synthesizing insights across multiple sources, verifying claims through cross-checking, and updating conclusions as new evidence emerges [2], [6]. While these ideas are still developing, they suggest a gradual shift toward AI systems that do more than simply retrieve information. Ideally, they begin to participate in the research process itself, assisting with tasks that traditionally required sustained human effort and careful reasoning.

III. RELATED WORK

Recent work on AI systems meant to assist with research tends to cluster around a few broad directions. One helpful way to think about the landscape is through three overlapping strands: retrieval-based approaches, reasoning-oriented language models, and the more recent movement toward agentic AI frameworks. Early efforts leaned heavily on retrieval-based improvements. The goal was fairly practical. Language models were powerful at generating text, yet they often lacked up-to-date or verifiable knowledge. Retrieval-Augmented Generation (RAG) emerged as a solution to this problem. Instead of relying only on what the model absorbed during training, the system first retrieves relevant documents and then produces a response grounded in those sources. In everyday use this can work surprisingly well. Ask about a particular dataset or a research technique, and the system may retrieve several relevant papers and summarize them with reasonable clarity. However, once the task shifts from answering questions to conducting deeper investigation, the weaknesses begin to appear. Multi-step reasoning, adapting when new evidence appears, or verifying whether earlier conclusions actually hold up are tasks these systems struggle with [1]. Part of the difficulty lies in the design itself. Most RAG pipelines operate in a single pass: retrieve, generate, and stop. Without persistent memory or mechanisms for planning what should happen next, their usefulness for complex research workflows remains somewhat limited [2].

Researchers soon began experimenting with ways to improve the reasoning behavior of language models. One strategy involved structured prompting techniques that encourage models to reason through intermediate steps rather than jumping directly to an answer. Methods such as Chain-of-Thought prompting, Least-to-Most prompting, and Tree-of-Thoughts follow this logic [10]–[12]. Instead of treating a question as a single block of computation, the model is guided to break the problem into smaller stages and process them sequentially. In many cases this noticeably improves logical consistency and problem-solving performance. Still, there is a limit to what prompting alone can achieve. These approaches remain tied to a static prompt structure. The reasoning occurs inside that structure, and once the output is generated the process effectively ends. The model does not independently decide to search for additional information, reconsider earlier assumptions, or test alternative explanations. When research tasks extend across multiple stages of exploration, that limitation becomes difficult to ignore [13].

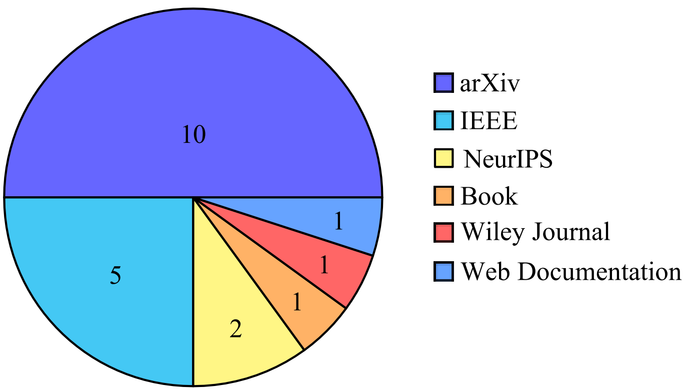

Fig. 1. Distribution of Publications by Source

More recent work has therefore turned toward the idea of agentic AI systems. The shift is partly conceptual. Instead of building systems that simply respond to queries, researchers are exploring architectures in which AI behaves more like an autonomous problem solver. In this framework, systems pursue goals, plan actions, interact with external tools, and adjust their behavior in response to feedback from their environment [3], [9]. Survey studies on agentic AI often describe architectures that integrate several components such as reasoning modules, memory systems, planning strategies, and mechanisms for executing external tools. When these elements are combined, the system can move through a sequence of investigative steps rather than producing a single response. These studies also attempt to clarify the distinctions between classical AI systems, simple task-oriented agents, and fully agentic frameworks capable of sustained autonomous activity. In doing so, they begin to outline the conceptual foundations for autonomous research engines [2].

An interesting development within this area is the emergence of multi-agent frameworks. Instead of relying on one large system to perform every task, several specialized agents collaborate within a shared environment. One agent might focus on retrieving documents, another on reasoning through evidence, while a third evaluates whether the conclusions actually make sense. Platforms such as AutoGen, MetaGPT, and LangGraph provide practical examples of how this coordination can work [4], [5], [14]. When agents divide responsibilities in this way, task decomposition becomes more manageable. Parallel exploration of multiple information sources also becomes possible, which is especially valuable when dealing with complex research questions that require comparing evidence across many documents. Iterative refinement is another advantage, since agents can critique or revise each other's outputs as the process unfolds [6].

Even with these promising developments, several challenges remain unresolved. Researchers frequently point to issues such as error propagation during multi-step reasoning, persistent hallucination problems, and the difficulty of evaluating systems that produce long chains of intermediate reasoning. Questions about transparency and ethical responsibility also arise when autonomous systems begin making decisions that are not fully interpretable to users [2], [3]. While many studies discuss the broader applications and social implications of agentic AI, frameworks specifically designed for deep research automation are still relatively uncommon [9].

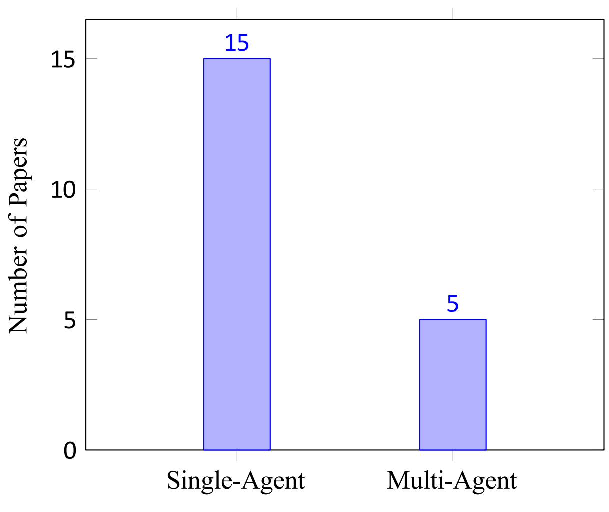

Fig. 2. Distribution of Papers by Agent Architecture

IV. SYSTEM ARCHITECTURE

The proposed system architecture brings together reasoning, retrieval, planning, and verification within an agentic framework intended to support independent, goal-driven research. Rather than relying on the familiar linear structure used in conventional RAG pipelines, where information is simply retrieved and then used to generate a response, the architecture adopts a modular and layered design. This structure allows the research process to unfold iteratively. Intermediate findings can influence later stages, and the system can adapt its strategy as new information emerges [1], [2]. In contrast to the traditional retrieve–generate pattern, this approach attempts to make AI-assisted research more exploratory and adaptive.

To support this behavior, the architecture is organized into four main layers: the Knowledge and Tool Layer, the Evaluation and Feedback Layer, the Agent Orchestration Layer, and the User Interaction Layer. A layered abstraction like this may seem familiar from other areas of system design, but it also aligns with architectural patterns discussed in recent work on agentic and autonomous AI systems. Researchers increasingly argue that separating reasoning, planning, tool usage, and evaluation into distinct components makes it easier to build systems that remain flexible as they evolve [3], [9].

Research tasks first enter the system through the User Interaction Layer. Unlike conventional interfaces that expect short factual queries, this layer is designed to accept broader research objectives. A user might, for instance, ask the system to explore trends in reinforcement learning or analyze recent developments in medical imaging. The responsibility of this layer is to interpret such intentions and convert them into structured goals that can guide the underlying agentic processes. By doing so, the interaction moves beyond simple question answering and begins to resemble the initiation of a research workflow [2].

At the core of the architecture lies the Agent Orchestration Layer. This layer coordinates a collection of specialized agents responsible for different aspects of the research process. Some agents focus on planning how the investigation should proceed, others retrieve relevant sources, while additional agents handle reasoning, synthesis, or verification tasks. A central coordinator manages the overall execution flow by breaking large research goals into smaller subtasks and assigning them to the appropriate agents. These agents may operate sequentially when tasks depend on earlier results, or concurrently when multiple research directions can be explored in parallel. Another important feature at this level is the presence of memory components that retain intermediate findings, contextual information, and previous decisions across multiple iterations. Such memory structures are widely regarded as essential for enabling sustained reasoning within agentic and multi-agent systems [4], [5].

The Knowledge and Tool Layer provides access to external resources that support deeper analysis. This includes web search interfaces, scholarly databases, document repositories, and various analytical tools. Through retrieval mechanisms, the system can dynamically gather relevant information rather than relying solely on a fixed internal knowledge base. At the same time, tool-calling capabilities allow agents to perform structured operations such as summarizing documents, comparing evidence from multiple sources, extracting citations,

TABLE I
SUMMARY OF LITERATURE REVIEW AND IDENTIFIED RESEARCH GAPS

Sr. No.	Research Paper Title (Year)	Methodology	Gap / Limitations
1	AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation (2023) [4]	Task breakdown and cooperation are made possible via a multi-agent conversational framework.	Inadequate self-verification and poor management of contradicting data.
2	MetaGPT: Meta Programming for Multi-Agent Collaborative Framework (2023) [5], [15]	Coordinated role-based multi-agent software development framework.	Lacks adaptive thinking and real-time information discovery; depends on predetermined roles.
3	ReAct: Synergizing Reasoning and Acting in Language Models (2023) [13]	Improves job execution by combining tool usage with reasoning trails.	There is little long-term planning and little overt multi-agent cooperation.
4	Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks (2020) [1]	Enables text creation by retrieving documents from static vector databases.	Lack of self-correction mechanisms, superficial thinking, and a static knowledge base.
5	Agentic AI: Autonomous Intelligence for Complex Goals (2025) [3]	An overview of planning, memory, and autonomy in agentic AI systems.	Mostly conceptual; deep research challenges need implementation-level validation.
6	Learning to Use Tools via Cooperative and Interactive Agents with Large Language Models (2024) [10]	Through organized inter-agent communication and action distillation, a cooperative multi-agent framework (ConAgents) with specialized agents for tool grounding, execution, and evaluation is made possible.	High coordination complexity, higher computing overhead, and less assessment of RAG procedures and large-scale real-world research.
7	AutoGPT: An Autonomous GPT-Based Agent (2023) [16]	An autonomous agent framework that links memory, tool utilization, and LLM reasoning to plan and carry out activities iteratively in the direction of user-defined objectives.	Weak self-verification, unstable long-horizon behavior, and a dearth of organized multi-agent coordination and assessment benchmarks.
8	LangGraph: Stateful Agentic Workflows (2024) [14]	Graph-based framework for creating multi-agent, stateful workflows with clear control over execution flow and agent transitions.	It is mostly an orchestration framework and does not deal with autonomous research assessment, retrieval optimization, or reasoning quality.
9	ReAct: Synergizing Reasoning and Acting in Language Models (2023) [13]	Chain-of-thought reasoning and external tool interaction are combined in a single-agent setting using the thinking-and-acting paradigm.	There is little long-term planning, no retrieval-grounded verification, and no explicit multi-agent cooperation.
10	MetaGPT: Meta Programming for Multi-Agent Collaborative Framework (2023) [12]	This role-based multi-agent software development framework makes use of predefined agent roles and structured communication protocols.	relies on static role definitions and processes rather than dynamic agent coordination, adaptive reasoning, or real-time knowledge retrieval.
11	Self-Refine: Iterative Refinement with Large Language Models (2023) [6]	Single-model iterative refinement method in which the LLM uses self-feedback loops to produce, evaluate, and enhance its own outputs.	Limited scalability for intricate research workflows; single-agent configuration without retrieval grounding, tool integration, or multi-agent cooperation.
12	A Collaborative Multi-Agent Approach to Retrieval-Augmented Generation Across Diverse Data Sources (2024) [11]	Multi-agent RAG architecture with modular query execution that uses specialized agents for various data sources (relational, document, NoSQL).	Limited self-reflection and autonomous planning; no unified agent orchestration or long-horizon thinking.
13	Generative to Agentic AI: Survey, Conceptualization, and Challenges (2025) [2]	With a focus on reasoning, planning, memory, and interaction, this conceptual and comparative study examines the transition from generative AI to agentic AI.	Mostly theoretical; neither executable architectures nor empirical support for agentic research systems are suggested.
14	Agentic AI in Education: State of the Art and Future Directions (2025) [17]	Domain-specific survey that categorizes agentic AI systems according to their functions, autonomy, and uses in educational settings.	The findings are not immediately applicable to general-purpose agentic RAG research engines; the application scope is restricted to schooling.
15	Agentic AI: Autonomous Intelligence for Complex Goals—A Comprehensive Survey (2025) [18]	Thorough examination of agentic AI applications, planning methods, architectures, foundations, and ethical issues.	Lacks implementation-level advice and benchmarking for multi-agent RAG systems; instead, it focuses on taxonomy and principles.
16	Agentic AI: A Comprehensive Survey of Technologies, Applications, and Societal Implications (2025) [9]	Comprehensive analysis of agentic AI technology, encompassing multi-agent coordination, autonomy, reinforcement learning, and social effects.	Does not specify an empirically verified agentic RAG pipeline or a single operational paradigm for deep research projects.

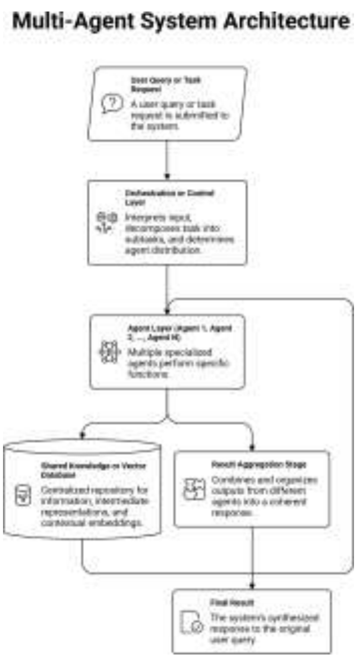

Fig. 3. Multi-Agent General Architecture

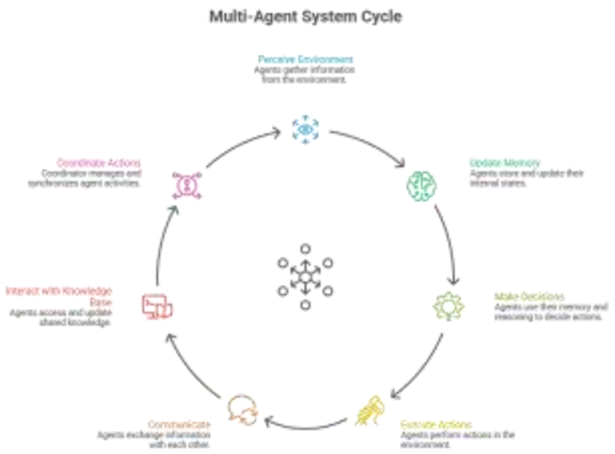

Fig. 4. Multi-Agent Systems cycle

or validating claims across references. Systems that combine reasoning with external tools tend to produce more grounded and informative outputs, particularly when working with large or complex information environments [13], [14].

Another important component is the Evaluation and Feedback Layer, which introduces a mechanism for self-reflection within the research process. Outputs generated during earlier stages are assessed for consistency, relevance, and factual accuracy [19]. The results of this evaluation are then fed back into the orchestration process, allowing agents to revisit earlier steps, refine their analysis, or seek additional evidence if necessary. This iterative feedback loop acts as a form of self-correction, helping reduce hallucinated information and improving the reliability of generated research outputs. Similar strategies of reflective reasoning and iterative refinement have been explored in recent work on agentic and reasoning-based AI systems [6], [9].

Overall, the architecture emphasizes modularity, autonomy, and explainability. By combining agent-based coordination with retrieval, reasoning, and evaluation mechanisms, the system moves beyond the limitations of static or purely reactive research assistance tools. The broader intention is to enable AI-assisted research processes that are not only scalable but also capable of adapting and improving as the investigation unfolds [2], [3].

V. RESULTS AND ANALYSIS

The performance of the proposed agentic research process can be better understood when placed beside more traditional Retrieval-Augmented Generation (RAG) systems. The comparison here is not based on benchmark scores or tightly controlled numerical evaluations. Instead, the analysis focuses on observable behavioral differences, architectural advantages, and the qualitative nature of the outputs produced by each approach. In other words, the question is less about raw optimization and more about how the systems actually behave during research-oriented tasks [1], [2].

1. Research Depth and Quality of Outputs

Traditional RAG systems often generate responses that remain relatively surface-level. They typically retrieve a limited set of documents from a pre-indexed vector database and summarize the most relevant passages [1]. While this can work well for straightforward questions, the process rarely moves beyond summarization. An agentic research process, however, approaches the task differently. Instead of treating the problem as a single retrieval step, the system decomposes the research objective into smaller tasks handled by specialized agents. Planning agents outline the investigation, retrieval agents gather sources, and reasoning agents analyze the relationships between those sources. Verification components then examine whether the conclusions actually follow from the evidence. Through this layered interaction, the system becomes capable of comparing multiple documents, explaining contradictions between sources, and synthesizing nuanced ideas. The resulting outputs tend to appear more structured and closer to what one might expect from a carefully written research summary, aligning with observations reported in recent studies on agentic and multi-agent AI systems [3], [9].

2. Information Freshness and Coverage

Another limitation of conventional RAG pipelines is their reliance on pre-indexed knowledge stores. Once the vector database is built, the system typically retrieves only from that fixed collection. If the database is outdated or incomplete, the resulting answers reflect those limitations [1]. An agentic approach attempts to mitigate this problem through iterative retrieval and autonomous information discovery. Rather than stopping after the first retrieval cycle, the system expands its search space as new leads appear during the reasoning process. A reference mentioned in one paper may trigger additional searches, which in turn introduce new sources into the analysis. This iterative exploration tends to broaden coverage and improve relevance, particularly in domains where information changes quickly, such as technology policy or financial analysis [2], [3].

3. Reduction of Hallucinations and Factual Errors

Hallucinated information remains one of the more persistent challenges in generative AI systems [9]. In single-pass generation pipelines, once an incorrect assumption appears in the output, there is little opportunity for correction. An agentic framework addresses this problem through repeated verification cycles. Claims generated during reasoning stages can be cross-checked against multiple sources within a shared knowledge store. If inconsistencies appear, verification agents flag them and prompt the system to re-examine the evidence. This process resembles a simplified form of peer review within the architecture itself. Over several iterations, unsupported claims are more likely to be identified and revised, which reduces the frequency of fabricated details compared to one-shot generation methods [6], [13].

4. Reasoning Transparency and Traceability

Another interesting difference concerns transparency. In many traditional RAG pipelines, the reasoning behind a final answer is largely hidden. The system retrieves documents, generates text, and presents the result with little explanation of how intermediate conclusions were reached [1]. An agentic architecture, by contrast, often records reasoning traces and confidence annotations within a shared knowledge structure. These traces capture intermediate decisions, evidence used during analysis, and the steps taken to arrive at a conclusion. The presence of such records makes it possible to review or audit the reasoning process afterward. From a research perspective, this traceability is valuable because it improves interpretability and allows users to understand how conclusions emerged from the available evidence [2].

5. Efficiency and Task Scalability

At first glance, agentic systems may appear computationally heavier because they involve coordination between multiple agents. There is indeed some orchestration overhead. Yet when applied to complex, long-horizon research tasks, the overall workflow can become more efficient. Retrieval, reasoning, and verification processes can run in parallel rather than sequentially. Different agents may explore separate research directions at the same time, and their findings can later be combined. Compared with manual workflows or single-agent pipelines, this parallel exploration often shortens the time required to synthesize large volumes of information. Similar scalability advantages have been observed in recent multi-agent frameworks [4], [5].

6. Comparative Summary

Taken together, these observations suggest that agentic research architectures offer several advantages over conventional RAG-based systems. Improvements appear particularly noticeable in areas such as research depth, factual reliability, adaptability, and reasoning transparency [2], [3]. Traditional RAG pipelines remain useful for quick factual queries or short informational summaries. However, when the objective shifts toward deeper research tasks that require reasoning, validation, and iterative refinement, agentic approaches begin to show clearer benefits.

VI. CHALLENGES AND LIMITATIONS

Agentic research engines offer several appealing advantages over conventional RAG-based systems, yet they also introduce a range of conceptual, operational, and technological challenges. Understanding these limitations is important, not simply as a critique, but as a way to realistically assess where such systems currently stand and where future improvements might be needed. The discussion around agentic AI often focuses on its potential, but the practical constraints are equally significant [2], [3].

1. System Complexity and Orchestration Overhead

A noticeable challenge arises from the architectural complexity of multi-agent systems. Compared with a traditional RAG pipeline, which typically follows a straightforward retrieve-and-generate sequence, agentic architectures involve several interacting components [1]. Tasks must be decomposed, agents must coordinate their activities, and communication between agents has to be carefully managed. If this coordination is poorly designed, the system may fall into inefficient cycles such as repeated execution of the same steps or reasoning loops that do not converge. Researchers working on multi-agent frameworks have noted that such orchestration problems can significantly increase development effort and make debugging far more difficult than in simpler AI pipelines [4], [5].

2. Computational and Resource Cost

Another concern relates to computational demands. Agentic systems often rely on iterative reasoning cycles and may run multiple agents simultaneously while exploring different lines of inquiry. This parallelism can improve the depth and quality of analysis, yet it also introduces higher computational overhead. Energy consumption, latency, and infrastructure costs tend to increase when compared with more lightweight retrieval-based systems [9]. In practice, these requirements may limit deployment in environments with restricted computational resources or in applications where rapid real-time responses are necessary [3].

3. Evaluation and Benchmarking Difficulties

Evaluating agentic research systems remains an open problem. Traditional metrics used in question-answering tasks, such as accuracy scores or BLEU-based comparisons, do not fully capture qualities like reasoning depth, evidence synthesis, or the ability to reconcile conflicting sources [2]. A system might generate a well-structured research summary while still scoring poorly on conventional benchmarks, or the opposite could occur. Because standardized evaluation frameworks for agentic research workflows are still emerging, comparing results across studies becomes challenging. This lack of common benchmarks also complicates efforts to reproduce experimental findings [9].

4. Error Propagation Across Agents

Although verification mechanisms can reduce hallucinations, errors may still propagate through the system when multiple agents are involved. If a retrieval agent gathers incomplete or biased information, the reasoning and synthesis agents that rely on that data may unknowingly incorporate those flaws into their conclusions. Once such assumptions enter the workflow, they can influence several later steps unless explicitly detected and corrected. For this reason, many researchers emphasize the need for stronger error isolation mechanisms and confidence-aware decision strategies within agentic pipelines [6], [13].

5. Trust, Transparency, and Explainability

Transparency is another area where challenges remain. Some agentic architectures attempt to record reasoning traces or intermediate decisions, which helps users understand how certain conclusions were reached. Even so, explaining the behavior of large language model–driven agents remains difficult. The internal reasoning processes of LLMs are often opaque, and the explanations produced by the system may themselves be approximations rather than precise reflections of the underlying computation. In research domains where decisions carry significant consequences, such as healthcare or financial analysis, this limited interpretability can raise concerns about reliability and accountability [2], [3].

6. Ethical and Governance Concerns

The autonomy of agentic research systems also raises ethical questions. When systems independently retrieve, synthesize, and evaluate information, the reliability of their sources becomes critical. Without proper safeguards, agents may prioritize highly visible or persuasive content rather than carefully validated evidence [9]. Bias present in the training data or retrieval sources may also be amplified during synthesis. Moreover, excessive autonomy without appropriate human oversight introduces the possibility of misuse or unintended outcomes, especially when such systems are applied in sensitive decision-making contexts [2] [20].

7. Scope Limitations and Non-AGI Nature

Finally, it is important to recognize that agentic research engines, despite exhibiting behaviors such as planning, reflection, and iterative correction, remain fundamentally limited systems. They operate within predefined task boundaries and depend heavily on the tools, data sources, and models available to them [3]. Although they may appear to demonstrate capabilities associated with broader artificial intelligence, they do not possess general reasoning ability, independent goal formation, or the kind of flexible common-sense understanding associated with artificial general intelligence [7]. In practice, their effectiveness is tied to well-defined research objectives and controlled environments rather than unrestricted problem solving.

VII. FUTURE SCOPE & CONCLUSION

This review explored the emerging idea of agentic research engines as a step forward from traditional Retrieval-Augmented Generation (RAG) systems. RAG architectures have certainly improved how information can be accessed and summarized by combining retrieval with language generation. That improvement is real. Yet the approach still carries some clear limitations. Many RAG systems rely on static knowledge bases and relatively shallow reasoning steps, and they often struggle to adapt when the information landscape shifts quickly. Because of this, their usefulness starts to fade when the task moves beyond simple question answering and into more demanding research work. Tasks that involve iterative exploration, comparison across multiple sources, or careful verification of conclusions often expose these weaknesses.

Looking through recent literature, one can see why interest in agentic AI architectures has grown. These systems introduce autonomous workflows built around multiple specialized agents that cooperate during the research process. Instead of simply retrieving documents and summarizing them, such architectures include components responsible for planning tasks, gathering sources, reasoning through evidence, and verifying conclusions. Shared knowledge stores and feedback loops allow the system to revisit earlier steps and refine its outputs. In effect, the design begins to resemble the way a human researcher might approach a problem: gather evidence, reconsider assumptions, and occasionally revise conclusions. That shift from passive retrieval toward active reasoning opens the possibility of deeper analysis, fresher information coverage, and clearer reasoning traces. Still, the approach is not without its difficulties. Questions around evaluation methods, computational cost, scalability, and governance remain unresolved and deserve careful attention [2], [3].

Looking ahead, several research directions may further strengthen the capabilities of agentic research systems.

1. Adaptive Agent Coordination and Task Decomposition

One area that seems particularly promising concerns how agents coordinate their work. Current implementations often depend on predefined orchestration strategies. Those strategies work reasonably well for predictable tasks, although real research rarely follows such tidy patterns. Future systems might benefit from adaptive coordination mechanisms that can break down complex research questions into smaller subtasks on the fly. Learning-based orchestration strategies could allow systems to decide which agents should collaborate, how tasks should be sequenced, and how deeply a particular problem should be explored. In principle, this would allow the workflow itself to adjust depending on the complexity of the research objective [4], [5].

2. Real-Time Knowledge Acquisition from Live Web Sources

Another direction involves strengthening access to real-time information. Many current research systems still rely on static document collections or pre-indexed databases. The problem is obvious. Knowledge in many fields changes quickly. Integrating dynamic web retrieval mechanisms, continuous knowledge updates, and stronger source validation processes could help ensure that research outputs reflect current developments rather than outdated snapshots of information. Such improvements may be especially important in fields like technology policy, economics, or biomedical research, where new findings appear regularly [2], [9].

3. Deeper Document Intelligence for Research Materials

Research documents themselves introduce another challenge. Academic papers, technical reports, and policy documents are rarely simple blocks of text. They contain figures, tables, equations, and terminology that can confuse basic text-processing systems. Advancing document intelligence therefore remains an important direction. Systems that develop stronger semantic understanding of PDFs and technical manuscripts could extract structured insights, identify relationships across sections, and connect ideas across multiple documents. With such capabilities, agentic systems might move beyond simple summarization toward genuine synthesis of research knowledge [2].

4. Multi-Source Reasoning and Conflict Detection

Another promising direction involves improving how systems handle disagreement between sources. Anyone who has spent time reading research papers knows that studies often contradict one another. Different datasets, assumptions, or experimental setups can lead to competing conclusions. Future agentic systems could incorporate mechanisms that detect such contradictions, compare supporting evidence, and assess the credibility of competing claims. Rather than simply aggregating information, the system would analyze disagreements and explain why they occur. That sort of analytical behavior would bring AI-assisted research a little closer to the reasoning process of human scholars [2], [3].

5. Iterative Self-Verification and Hallucination Reduction

Reducing hallucinated information continues to be one of the more persistent challenges in large language model–based systems. Retrieval improves factual grounding, yet incorrect claims can still appear. Future work may therefore emphasize stronger verification loops. In such workflows, agents repeatedly compare generated conclusions with retrieved evidence and refine their reasoning when inconsistencies appear. Self-correcting reasoning modules and verification agents could gradually improve factual accuracy and strengthen trust in AI-assisted research tools [3], [9].

6. Automated Research Report and Paper Generation

Another interesting direction involves the generation of structured research documents. Producing short summaries is relatively straightforward, but writing a coherent research-style report is far more demanding. Future architectures may focus on organizing arguments logically, synthesizing evidence from multiple sources, and acknowledging uncertainty when conclusions remain tentative. If done well, such systems could assist researchers not only in gathering information but also in drafting literature reviews and analytical reports that follow recognizable academic structures [2], [9].

7. Ethical Governance and Source Reliability Mechanisms

As these systems become more autonomous, questions of governance become harder to ignore. Autonomous research engines must evaluate the credibility of the sources they rely on. Future systems might incorporate credibility scoring for information sources, mechanisms for detecting bias, and policy-aware reasoning modules that flag potentially unreliable or misleading material. Such safeguards will be necessary if AI-assisted research tools are to remain trustworthy rather than becoming amplifiers of misinformation [2], [9].

In summary, agentic research engines suggest an interesting shift in how AI may support research activities. They move the focus away from passive information retrieval and toward more active forms of knowledge synthesis. Continued progress in adaptive coordination, real-time retrieval, document intelligence, verification mechanisms, and ethical governance will likely determine whether these systems mature into reliable research assistants capable of supporting complex knowledge discovery tasks.

REFERENCES

[1] P. Lewis *et al.*, "Retrieval-augmented generation for knowledge-intensive nlp tasks," *Advances in Neural Information Processing Systems*, 2020.

[2] J. Schneider, "Generative to agentic ai: Survey, conceptualization, and challenges," *arXiv preprint arXiv:2504.18875*, 2025.

[3] D. B. Acharya, K. Kuppan, and B. Divya, "Agentic ai: Autonomous intelligence for complex goals—a comprehensive survey," *IEEE Access*, 2025.

[4] Q. Wu *et al.*, "Autogen: Enabling next-gen llm applications via multi-agent conversation," *arXiv preprint arXiv:2308.08155*, 2023.

[5] S. Hong *et al.*, "Metagpt: Meta programming for multi-agent collaborative framework," *arXiv preprint arXiv:2308.00352*, 2023.

[6] A. Madaan *et al.*, "Self-refine: Iterative refinement with large language models," *arXiv preprint arXiv:2303.17651*, 2023.

[7] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. Pearson, 2021.

[8] V. Ingale, B. Wankar, K. Jadhav, T. Adedoja, V. K. Borate, and Y. K. Mali, "Healthcare is being revolutionized by ai-powered solutions and technological integration for easily accessible and efficient medical care," in *2024 15th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. IEEE, 2024, pp. 1–6.

[9] A. K. Pati, "Agentic ai: A comprehensive survey of technologies, applications, and societal implications," *IEEE Access*, 2025.

[10] J. Wei *et al.*, "Chain-of-thought prompting elicits reasoning in large language models," *Advances in Neural Information Processing Systems*, 2022.

[11] D. Zhou *et al.*, "Least-to-most prompting enables complex reasoning in large language models," *arXiv preprint arXiv:2205.10625*, 2022.

[12] S. Yao *et al.*, "Tree of thoughts: Deliberate problem solving with large language models," *arXiv preprint arXiv:2305.10601*, 2023.

[13] S. Yao, J. Zhao *et al.*, "React: Synergizing reasoning and acting in language models," *arXiv preprint arXiv:2210.03629*, 2023.

[14] LangChain Team, "Langgraph: Stateful agentic workflows," https://langchain-ai.github.io/langgraph/, 2024.

[15] X. L. Dong, E. Gabrilovich, K. Murphy, V. Dang, W. Horn, C. Lugaresi, S. Sun, and W. Zhang, "Knowledge-based trust: Estimating the trustworthiness of web sources," *arXiv preprint arXiv:1502.03519*, 2015. [Online]. Available: https://arxiv.org/abs/1502.03519

[16] T. B. Richards, "Autogpt: An autonomous gpt-based agent," https://github.com/Significant-Gravitas/AutoGPT, 2023.

[17] G. Kostopoulos, V. Gkamas, M. Rigou, and S. Kotsiantis, "Agentic ai in education: State of the art and future directions," *IEEE Access*, 2025.

[18] X. Chen *et al.*, "Plan-and-execute: An agentic framework for task completion," *arXiv preprint*, 2023.

[19] S. Murugesan, "The rise of agentic ai: Implications, concerns, and the path forward," *IEEE Intelligent Systems*, vol. 40, no. 2, pp. 8–14, 2025.

20. Y. K. Dwivedi, M. Y. Helal, I. A. Elgendy, R. Alahmad, P. Walton, A. Suh, V. Singh, and I. Jeon, "Agentic ai systems: What it is and isn't," *Global Business and Organizational Excellence*, vol. 45, no. 3, pp. 253–263, 2026.