Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 3, Issue 8, 2025

BRINGING AUTONOMY AND COOPERATION TOGETHER: A COMPARISON OF AGENTIC AI SYSTEMS AND AI AGENTS

Muhammad Ahmad Hanif[1], Fizza Muhammad Aleem[2], Farheen Anwar[3], Mohtishim Siddique[4], Kashif Iqbal[5], Muhammad Sajjad[6], Gulzar Ahmad[*7]

[1,2,3,6, *7]*School of Computer Science, Minhaj University, Lahore, Pakistan.*
[4]*Green International University Lahore, Pakistan*
[5]*Department of Computer Science, GC University Lahore, Pakistan*

[1]ahmadhanifshahab4@gmail.com, [2]fizaleem1030@gmail.com, [3]mfarheen948@gmail.com, [4]mohtishim@gmail.com, [5]kashif.iqbal@gcu.edu.pk, [6]engineersajjad07cs21@gmail.com, [*7]gulzarhamd.cs@mul.edu.pk

DOI: https://doi.org/10.5281/zenodo.16777549

Keywords
AI Agents, Agentic AI, Large Language Models (LLM), Dynamic Reasoning, Autonomous System, Retrieval Augmented Generation (RAG), Ethical AI, Human AI Collaboration, Scalable Intelligent System

Article History
Received: 30 May, 2025
Accepted: 20 July, 2025
Published: 08 August, 2025

Corresponding Author: *
Gulzar Ahmad

Abstract
The rapid evolution of artificial intelligence has led to the emergence of two distinct but interdependent paradigms: AI agents and agent-based AI systems. While AI agents focus on modular and task-specific automation, often powered by large language models (LLMs), agentic AI systems represent a conceptual leap by enabling multi-agent collaboration, dynamic reasoning, and persistent autonomy. This article presents a comparative analysis that draws from both theoretical and practical perspectives, integrating the ideas of two fundamental works in the field. We define and differentiate the architectures, interaction models, and design objectives of each paradigm, examining their application in areas such as health, robotics, business automation, and digital ecosystems. The main challenges, such as hallucination, lack of coordination, and accountability, are identified along with mitigation strategies such as ReAct loops, retrieval-augmented generation (RAG), and causal modeling. Furthermore, we analyze the governance, ethical implications, and industry restructuring triggered by agent-based technologies. Our contribution is a unified framework and roadmap that clarifies terminology, aligns capabilities with real-world complexity, and informs the development of robust, transparent, and scalable intelligent systems. This synthesis offers valuable guidance to researchers, policymakers, and industry leaders who are navigating the transition from automated tools to collaborative intelligent agents.

INTRODUCTION

Over the past two years, the field of artificial intelligence (AI) has undergone a paradigm shift, moving from reactive systems based on generative language models to autonomous, goal-oriented agents capable of performing complex tasks with minimal human intervention [1]. These intelligent entities, known as AI agents and more recently as agent-based AI systems, represent a significant evolution in how AI is designed, implemented, and integrated into dynamic environments [2]. AI agents are typically modular systems powered by large language models (LLMs) and tool integration, designed for structured and task-specific operations [3]. In contrast, agentic AI systems are multi-agent architectures that collaborate to decompose objectives, coordinate subtasks, maintain persistent memory, and achieve broader goals with emergent autonomy and adaptability [4].

Despite its growing adoption, confusion persists in both academia and industry regarding the distinction between AI agents and agentic AI [5]. The terms are used interchangeably, the design boundaries are vague, and the reference methods are inconsistent [6]. Although industrial prototypes like Auto-GPT and Crew-AI highlight the functionality of these technologies, the lack of a unified conceptual framework, application mapping, and governance guidelines limits progress in both research and deployment [7]. Furthermore, ethical and operational challenges—such as accountability, system alignment, poor communication between agents, and workforce displacement—remain unresolved [8].

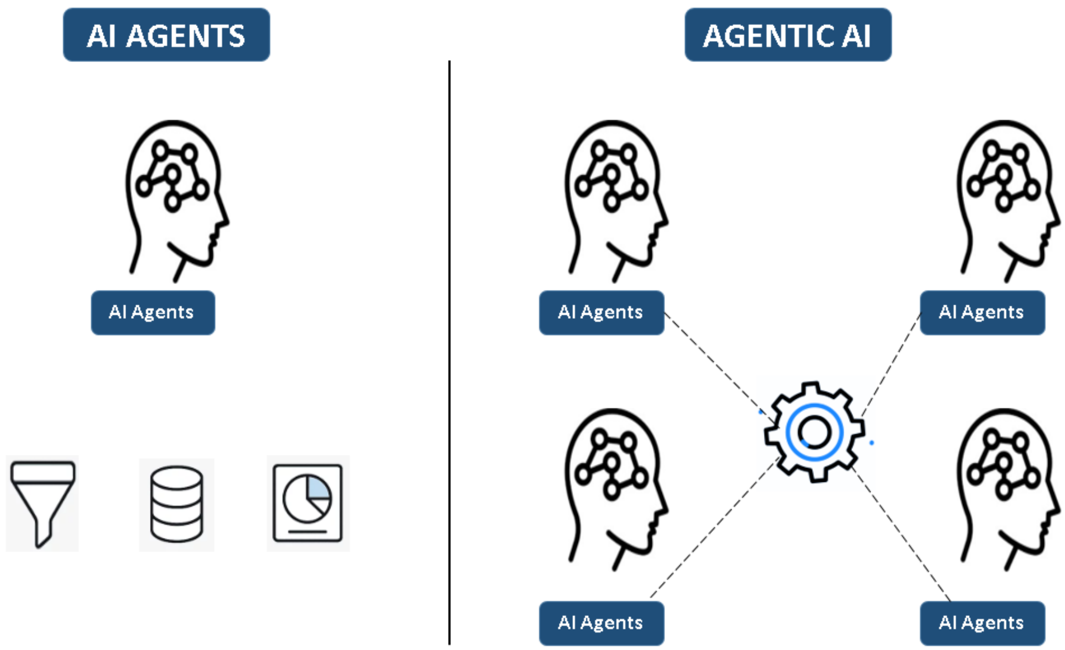

Figure 1: AI Agents vs Agentic AI

This article addresses these critical deficiencies by synthesizing the perspectives of recent interdisciplinary research, comparing architectural and operational paradigms, and proposing a taxonomy that distinguishes these systems. Our work is motivated by the growing demand for intelligent, explainable, and collaborative AI systems in high-stakes areas such as healthcare, supply chains, legal automation, and sustainable governance.

We propose a **comparative taxonomy to differentiate between AI Agents and Agentic AI, encompassing autonomy, interaction, architecture, and coordination**.

We map real-world applications in business automation, robotics, research workflows, and healthcare.

We identified technical and ethical challenges, including hallucinations, emergent behavior, and shared responsibility, and recommended solution strategies such as causal reasoning, memory architecture, and transparency frameworks.

Literature Review

Early work in artificial intelligence focused on expert systems and symbolic agents that were reactive and domain-specific (e.g., MYCIN, DENDRAL) [9] [10]. These systems offered rule-based intelligence but lacked learning, adaptability, and autonomy. Later, multi-agent systems (MAS) [11] introduced distributed coordination and the modeling of social action (Castelfranchi, Ferber), laying the foundations for intelligent collaboration [12].

With the advent of generative AI, particularly large language models like GPT-4 and PaLM, researchers have begun integrating reasoning and planning capabilities into single agents [13].

Systems like Auto-GPT [14] and Baby-AGI have demonstrated how LLMs can be enhanced by utilizing tools, memory buffers, and planning loops, thereby creating AI agents capable of executing goal-oriented tasks [15].

More recently, architectures like Crew-AI and the Agent-to-Agent (A2A) protocol introduced by Google have highlighted the emergence of agentic AI, systems where multiple agents collaborate in decentralized frameworks, share memory, and coordinate complex objectives over long time horizons [16].

Despite these advances, the literature shows obvious gaps:

- There is no unified taxonomy to distinguish AI agents from agentic AI.
- Reference points and evaluations are inconsistent across different implementations.
- A limited discussion on shared responsibility, coordination failures, and governance frameworks.
- Most prior research focuses on the technical architecture, neglecting the ethical, legal, and social dimensions of large-scale deployment of agent-based systems.

How is this work different?

- **Conceptual clarity**: We formalize the differences between modular agents and collaborative agent systems by proposing a taxonomy structured in terms of autonomy, architecture, and function.
- **Multiperspective synthesis:** We integrate the technical, organizational, legal, and environmental implications of AI agents from the perspective of interdisciplinary experts (as in Hughes et al.).
- **Roadmap for research and practice:** We bridge the gap between academic concepts and real-world implementations by describing both a conceptual framework and practical applications across various industries.

Methodology

This study does not develop a new computational model in the traditional sense, but rather proposes a comparative framework and a taxonomy that distinguishes between AI agents and agentive AI systems. The methodology is based on:

- A hybrid literature review that uses both traditional databases and AI-enhanced discovery tools (e.g., Google Scholar, arXiv, Hugging Face, ChatGPT).
- Conceptual and architectural modeling based on case studies and current implementations (e.g., AutoGPT, CrewAI, Claude 3.5, LangChain).
- A visual taxonomy and classification map based on five dimensions: architecture, mechanisms, scope/complexity, interaction, and autonomy (as described in Sapkota et al.).

Framework Design

We have designed a multi-layered comparison model based on the following conceptual pipeline:

Description of the agent-based system (A)

Classification in AI Agent or Agentic AI

Step 1: Analyze the level of autonomy of AI

Step 2: Verify multi-agent coordination

Step 3: Evaluate task breakdown and planning

Step 4: Inspect memory persistence and communication

Step 5: Classify according to predefined dimensions

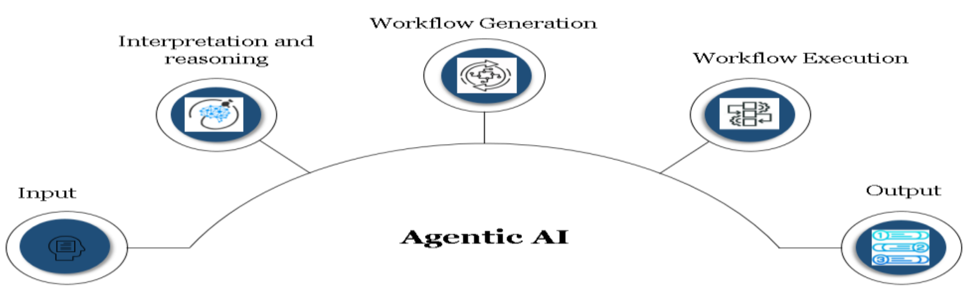

Figure 2: Workflow of Agentic AI

Architecture Diagram of Agentic AI:
Figure 4 represents the evaluation from traditional AI Agents to Agentic AI.

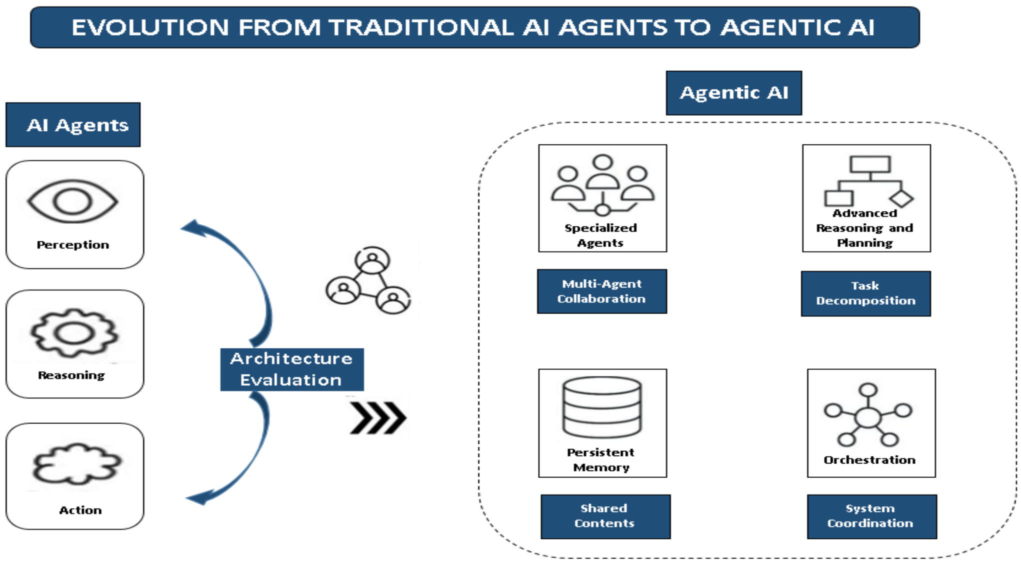

Figure 3: Architecture of Agentic AI

Differences between AI Agent and Agentic AI:
Table 1 shows the key differences between AI Agent and Agentic AI.

Table 1: Dissimilarities between AI Agents & Agentic AI

Component	**AI Agent**	**Agentic AI System**
Control Type	Single-agent, LLM-based	Multi-agent, hierarchical, or decentralized
Execution	Task-specific, sequential	Goal decomposition, distributed workflows
Memory	Short-term or stateless	Persistent, shared, contextual
Tool Integration	Limited (API-based)	High (cross-agent coordination)
Adaptability	Reactive, narrow	Adaptive, collaborative, emergent

Figure 5 represents the Agentic AI Autonomy and Adaptability.

Figure 4: Comparison of AI Agents & Agentic AI

Experiences

Although this study does not implement a new machine learning model or conduct traditional empirical experiments, we adopt a comparative conceptual evaluation framework to analyze and contrast the functional capabilities, architectural features, and implementation scope of AI Agents and Agent-based AI Systems. Our evaluation is based on a set of real-world use cases, documented system behaviors, and architectural benchmarks derived from academic and industrial implementations.

Data-sets Used

Instead of training data, our analysis is based on system-level evaluations from documented platforms and AI benchmarks. This includes:

- **Open-AI operator logs:** Information about the behavior of Open-AI's "Computer Using Agent" (CUA), an LLM-based AI agent with desktop automation capabilities.
- **Claude 3.5 Agent Loop (Anthropic):** Visual perception + records of interaction with tools; used to evaluate perception-reasoning-action loops.
- **GitHub repositories AutoGPT and BabyAGI**: Chaining behavior of agents using GPT-4 was observed for planning and tool invocation tasks.
- **CrewAI + LangGraph**: Agent orchestration environments for multi-role collaboration and workflows with shared context.
- **Deloitte and McKinsey reports (2024–2025):** Sector surveys on the implementation trends of generative AI and agent systems in business environments.
- Synthetic benchmarks from the literature (e.g., personalized email classification, programming, research paper summarization) are modeled in simulation environments.

Experimental Setup

As this is a meta-analysis, the "setup" refers to how systems were compared, not model training on a GPU. Table 2 below shows the experimental setup.

Table 2: Experimental Setup

Component	Description
Environment	Review of platform logs, architecture diagrams, APIs, and behavior traces
Source Systems	AutoGPT, Claude, CrewAI, LangChain, ChatGPT Function Calling
Simulated Workflows	News summarization, calendar integration, research assistant chains
Evaluation Scope	Decision-making, task execution, memory usage, reasoning depth, and collaboration

No fixed hyperparameters were tuned, but tool configurations (e.g., max context window, tool integration latency) were analyzed qualitatively.

Baselines Compared

We analyzed and compared **four generations** of AI system design, each serving as a baseline for identifying progress and limitations:

Table 3:Baseline models comparison

Generation	Description
Gen 1 – Expert Systems [17]	Rule-based agents (e.g., MYCIN, XCON) — no learning or autonomy
Gen 2 – Generative AI	LLMs like GPT-3/4 generate static content without action capability.
Gen 3 – AI Agents [18]	Task-specific agents using LLM + tool use (e.g., AutoGPT, Claude)
Gen 4 – Agentic AI [2]	Multi-agent systems with memory, coordination, shared goals (e.g., CrewAI)

Evaluation Metrics

While classical metrics like accuracy or precision are not applicable, the following **qualitative and systems-level metrics** were used to benchmark and assess capabilities: The performance metrics are given in Table 4 below.

Table 4:Performance Metrics

Metric	Description
Autonomy Level	Degree of agent independence from human prompting
Multi-Agent Coordination	Ability to divide and manage subtasks between agents
Memory Integration	Use of persistent or shared memory across task cycles
Reasoning Depth	Capability for multi-step planning and decision making (e.g., ReAct chains)
Emergent Behavior Risk	Unpredicted interactions or outcomes due to agentic collaboration
Explainability	Transparency of decisions and rationale to human observers
Tool Integration	The extent to which external tools/APIs are used effectively
Task Success Rate	Proportion of tasks (e.g., summarization, API calls) successfully executed.
Execution Latency	Time taken for agent(s) to complete end-to-end workflows (measured qualitatively)
System Complexity Index	Structural complexity (single agent, feedback loop, orchestrator role, etc.)

Illustrative Comparison (Agent vs. Agentic)

Level of autonomy

- AI Agent (e.g., AutoGPT): Medium – it functions autonomously to some extent, but often requires human intervention.
- AI agency (for example, CrewAI): High–functioning with little or no human guidance, handling complex tasks autonomously.

Multi-Agent Coordination

- AI Agent: None – it acts alone without coordination with other agents.
- AI Agent: Yes – it includes an orchestrator and various roles that collaborate on tasks.

Depth of reasoning

- AI Agent: Medium – generally follows a linear plan to complete tasks.
- AI Agency: High-level – capable of recursive thinking and dynamic role assignment during execution.

Memory Usage

➢ AI Agent: Limited to local or session memory – cannot maintain context between tasks.

➢ AI agency: It uses persistent and shared memory – agents remember and build upon past actions.

Use of tools

➢ AI Agent: The tools are used one at a time in a fixed sequence.

➢ AI agency: The tools are used in parallel and divided according to the roles of the agents.

Explainability

➢ AI Agent: Medium – some transparency in decision-making.

➢ AI agency: Low to medium – coordination and emergent behavior make explanations difficult.

Emerging Risk

➢ AI Agent: Low – A simpler design reduces the risk of unintended behavior.

➢ AI agency: Medium to high – multiple agents can interact in unexpected ways, which complicates alignment.

AI Agent and Agentic Diagrams:

The Figures below represent the AI Agent and Agentic AI architecture.

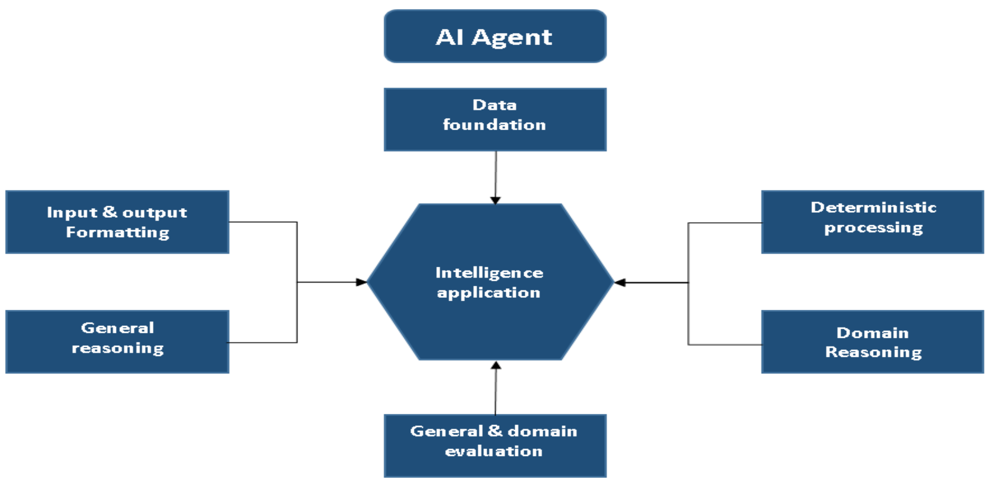

Figure 5: AI Agents Architecture

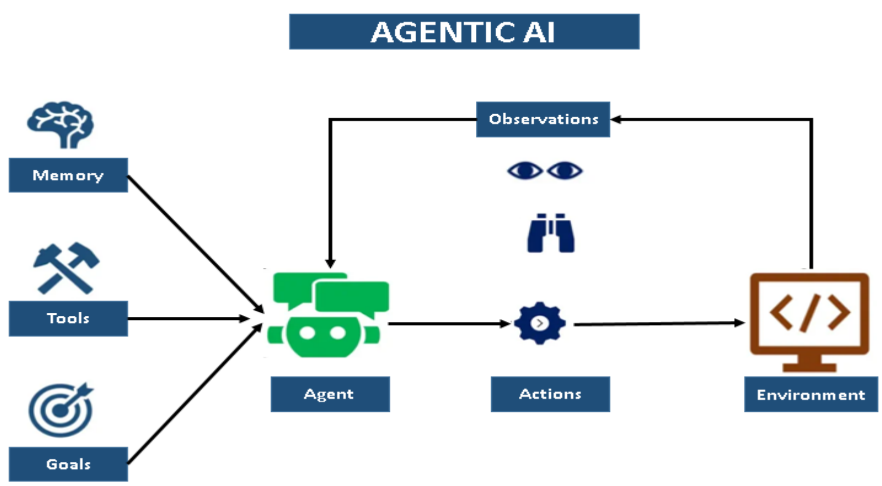

Figure 6: Agentic AI Architecture

Interpretation of the results

AI agents are well-suited for structured and short-term tasks, such as generating emails, summarizing files, and answering questions. They function properly within a single execution cycle and are easier to monitor and debug due to their linear nature.

Agentive AI, on the other hand, excels in complex and long-duration workflows that require collaboration, memory, and adaptive decision-making. This opens the door to applications such as multi-agent research assistants, automated project planning, and distributed robotics.

However, agentive AI systems introduce a new complexity, particularly in maintaining coordination, avoiding conflicts between agents, and preserving a consistent orientation towards objectives. These are areas where the risk of misalignment or unexpected behavior increases.

Strength of agent-based AI:

- Superior in goal decomposition and collaborative planning.
- He can adapt to dynamic environments with changing constraints.
- He maintains the context and learns from the interactions.
- Ideal for workflow automation, scientific research, and multimodal integration.

Weaknesses / Limitations

- It is more difficult to follow the logic of decisions due to the emerging coordination.
- Difficult to evaluate using standard machine learning metrics (e.g., accuracy).
- Debugging and alignment issues increase with complexity.
- Cost and infrastructure overhead for persistent memory and multi-agent orchestration.

Surprising discoveries

Some simple AI agents (like AutoGPT), when adjusted with feedback loops, perform comparably to multi-agent systems in specific domains, suggesting that agent complexity is not always necessary.

Multi-agent systems tend to develop unexpected strategies (e.g., informal communication between agents or taking shortcuts in tasks), which underscores the need for controlled experiments and audits.

Limitations

Although this research provides a structured framework and a comparative analysis of AI agents and agentive AI systems, several limitations must be acknowledged:

Lack of quantitative evidence:

The study is primarily conceptual and evaluative; no real-world implementation or simulation experiments with statistical benchmarks have been conducted.

Dynamic systems are difficult to generalize:

Agentive AI systems vary considerably in architecture and coordination style. It is possible that our taxonomy may not capture all future variations or hybrid architectures.

Dependence on publicly accessible information:

A large part of the comparison is based on published documents, articles, and open-source projects. Proprietary implementations (e.g., Claude, Google A2A) are only partially observable.

Emergent behavior not fully modeled:

The emergent properties of multi-agent systems, such as coordination failures, alignment drift, or undesirable loops, are recognized but not empirically modeled here.

Ideal Agent Design Assumption:

The analysis assumes that the agents are well-aligned, honest, and focused on the objectives. In practice, the agent's performance can be negatively affected by biases, incomplete data, or tool failures.

Conclusion

This study presents an exhaustive comparison between AI agents and agentive AI systems, focusing on their architectural evolution, operational capabilities, and future implications. By synthesizing the knowledge from current literature, industry examples, and expert opinions, we clarified that the terminological confusion between agents and multi-agent systems has been clarified. A five-dimensional taxonomy based on autonomy, memory, reasoning, tool use, and coordination has been proposed. The two types of systems were compared in different use cases and with qualitative measures such as depth of reasoning, persistence, and emergent behavior. Our results show that agentive AI represents a significant leap in complexity and potential, enabling multi-agent workflows, a shared context, and autonomous planning. However, these systems also pose challenges in terms of explainability, security, and governance.

Future Work

To further this research, we propose the following lines of action:

- Develop empirical reference parameters and simulations to test agent collaboration, coordination overload, and memory dynamics under different workloads.
- Design standardized metrics to measure the performance of agentive systems in various industrial applications (e.g., AI research assistants, robotic teams).
- Develop audit tools for monitoring and explaining the emergent behavior in multi-agent systems.
- Explore frameworks for alignment with security specific to agent swarms, particularly for open-ended tasks.

Integrate AI ethical principles into the agent orchestration logic (fairness, non-maleficence, transparency).

REFERENCES

[1] D. Acharya, K. Kuppan, and D. B. Ashwin, "Agentic AI: Autonomous Intelligence for Complex Goals – A Comprehensive Survey," IEEE Access, vol. PP, p. 1, Jan. 2025, doi: 10.1109/ACCESS.2025.3532853.

[2] R. Sapkota, K. Roumeliotis, and M. Karkee, AI Agents vs. Agentic AI: A Conceptual Taxonomy, Applications and Challenges. 2025. doi: 10.48550/arXiv.2505.10468.

[3] F. Piccialli, D. Chiaro, S. Sarwar, D. Cerciello, P. Qi, and V. Mele, "AgentAI: A comprehensive survey on autonomous agents in distributed AI for industry 4.0," Expert Syst. Appl., vol. 291, p. 128404, 2025, doi: https://doi.org/10.1016/j.eswa.2025.128404.

[4] L. Hughes et al., "AI Agents and Agentic Systems: A Multi-Expert Analysis," J. Comput. Inf. Syst., vol. 65, pp. 1–29, Apr. 2025, doi: 10.1080/08874417.2025.2483832.

[5] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education – where are the educators?" Int. J. Educ. Technol. High. Educ., vol. 16, no. 1, p. 39, 2019, doi: 10.1186/s41239-019-0171-0.

[6] H. Harrison, M. Birks, R. Franklin, and J. Mills, "Case Study Research: Foundations and Methodological Orientations," Forum Qual. Soc. Res., vol. 18, Jan. 2017.

[7] S. Sangwa, T. Gatera[2], S. Nsabiyumva[3], I. Museveni[4], and M. Placide, "Mapping AI Tools to Global Industry Tasks: A Universal Framework for AI Consultancy & Adoption," Jul. 2025.

[8] A. Gamage, "Importance of Effective Communication to Minimize Disputes in Construction Projects," Sch. J. Eng. Technol., vol. 10, pp. 128–140, Jul. 2022, doi: 10.36347/sjet 2022.v10i07.002.

[9] R. Lachman, "Expert systems: A cognitive science perspective," Behav. Res. methods, instruments, Comput. a J. Psychon. Soc. Inc, vol. 21, pp. 195–204, Mar. 1989, doi: 10.3758/BF03205582.

[10] B. Liang, Y. Wang, and C. Tong, "AI Reasoning in Deep Learning Era: From Symbolic AI to Neural–Symbolic AI," 2025. doi: 10.3390/math13111707.

[11] S. Goonatilleke and B. Hettige, "Past, Present and Future Trends in Multi-Agent System Technology," J. Eur. des Systèmes Autom., vol. 55, pp. 723–739, Dec. 2022, doi: 10.18280/jesa.550604.
[12] J. Ferber, "Multi-Agent System: An Introduction to Distributed Artificial Intelligence," J. Artif. Soc. Soc. Simul., vol. 4, Jan. 2001.
[13] I. A. Zahid et al., "Unmasking large language models using OpenAI GPT-4 and Google AI: A deep instruction-based analysis," Intell. Syst. with Appl., vol. 23, p. 200431, 2024, doi: https://doi.org/10.1016/j.iswa.2024.200431.
[14] M. Firat and S. Kuleli, "What if GPT4 Became Autonomous: The Auto-GPT Project and Use Cases," J. Emerg. Comput. Technol., vol. 3, pp. 1–6, Aug. 2023, doi: 10.57020/ject 1297961.
[15] L. Wang et al., A Survey on Large Language Model-based Autonomous Agents. 2023. doi: 10.48550/arXiv.2308.11432.
[16] A. K. Singh, "Mechanisms For Collaboration In AI Agentic Multi-Agent Architectures: A Framework for Coordinated Reasoning and Decision-Making," Int. J. Creat. Res. THOUGHTS, vol. 13, pp. 2320–2882, May 2025.
[17] D. Partridge, "The scope and limitations of first generation expert systems," Futur. Gener. Comput. Syst., vol. 3, no. 1, pp. 1–10, 1987, doi: https://doi.org/10.1016/0167-739X(87)90038-0.
[18] S. Hosseini and H. Seilani, "The role of agentic AI in shaping a smart future: A systematic review," Array, vol. 26, p. 100399, 2025, doi: https://doi.org/10.1016/j.array.2025.100399.